

# Residual Entropy of the Mott Insulator with No Symmetry Broken

Fusayoshi J. OHKAWA<sup>\*)</sup>

*Division of Physics, Hokkaido University, Sapporo 060-0810, Japan*

(Received March 14, 2012; revised May 7, 2012)

The half-filled ground state of the Hubbard model on the hypercubic lattice in  $D$  dimensions is studied by the Kondo-lattice theory, which is none other than the  $1/D$  expansion theory, but within the constrained Hilbert subspace where no symmetry is allowed to be broken. A gap can open in the single-particle excitation spectrum if and only if the residual entropy or entropy at  $T = +0$  K is nonzero. The Mott insulator with no symmetry broken, if it is possible, is characterized by nonzero residual entropy or nonzero entropy at  $T = +0$  K. This conclusion is consistent with Brinkman and Rice's theory and the dynamical mean-field theory. According to the well-known argument based on the Bethe-ansatz solution, on the other hand, the half-filled ground state in one dimension is the Mott insulator although its residual entropy per unit cell is vanishing in the thermodynamic limit. Two possible explanations are given for the contradiction between the present paper and the well-known argument.

Subject Index: 345, 346, 397

## §1. Introduction

Mott localization or the Mott insulator is a longstanding crucial issue.<sup>1)</sup> Whether magnetism is itinerant-electron or local-moment magnetism is directly related to the physics of the Mott insulator. High-temperature (high- $T_c$ ) superconductivity occurs in the vicinity of the phase of a local-moment type of antiferromagnetic insulator.<sup>2)</sup> Thus, it is anticipated that the mechanism of high- $T_c$  superconductivity is closely related to the physics of the Mott insulator.<sup>3), 4), 5)</sup>

For example, we consider the Hubbard model at the absolute zero Kelvin. According to Hubbard's theory,<sup>6), 7)</sup> when the on-site repulsion  $U$  is as large as or larger than the bandwidth  $W$ , the band splits into the upper and lower Hubbard bands, i.e., the Hubbard gap opens. According to Gutzwiller's theory,<sup>8), 9), 10)</sup> when  $U \gtrsim W$  and electrons are almost half-filled, electrons are renormalized into heavy quasi-particles. According to a slave-boson theory,<sup>11)</sup> a midband appears at the chemical potential between the upper and lower Hubbard bands; the midband is none other than a band of heavy quasi-particles in Gutzwiller's theory, so that it is called the Gutzwiller band. According to Brinkman and Rice's theory,<sup>12)</sup> which is based on Gutzwiller's theory, if electrons are exactly half-filled, a metal-insulator transition occurs at a critical  $U_c \simeq W$ . In an insulating phase at  $U > U_c$ , the Gutzwiller band disappears and the Hubbard gap becomes a complete gap. According to these theories, the ground state is a metal for any  $U$  unless electrons are exactly half-filled and, therefore, the ground state can be the Mott insulator if and only if electrons are exactly half-filled and the onsite  $U$  is large enough.

---

<sup>\*)</sup> E-mail: fohkawa@phys.sci.hokudai.ac.jp

According to Brinkman and Rice's theory,<sup>12)</sup> the specific-heat coefficient diverges as the onsite  $U$  approaches the critical  $U_c$  from the metallic or nongapped phase; and no symmetry is broken in the insulating or gapped phase. A combination of these two properties means or at least implies that the residual entropy is nonzero in the gapped phase, although it is unlikely that the third law of thermodynamics is broken. Whether the residual entropy is zero or nonzero in the gapped phase is a crucial issue to be elucidated.

Either Hubbard's or Gutzwiller's theory is under single-site approximation (SSA). If all the single-site terms are rigorously considered by a theory, the theory is the supreme SSA ( $S^3A$ ) theory. In the limit of  $D \rightarrow +\infty$ , where  $D$  is the spatial dimensionality, multisite terms vanish except for certain types of conventional Weiss mean field,<sup>\*</sup>) which is a multisite effect. Thus, the  $S^3A$  is rigorous in the limit of  $D \rightarrow +\infty$ ,<sup>13), 14), 15), 16)</sup> but within the constrained Hilbert subspace where no symmetry is allowed to be broken. A problem of rigorously calculating all the single-site terms under  $S^3A$  can be substituted for a self-consistent Kondo problem,<sup>17), 18), 19), 20)</sup> in which the Anderson model to be solved has to be self-consistently determined with the single-site terms. A dynamical mean-field theory<sup>20), 21)</sup> (DMFT) or a dynamical coherent-potential approximation (DCPA) theory<sup>22)</sup> is none other than the  $S^3A$  theory.

Under DMFT or  $S^3A$ , a metal-insulator transition occurs at  $U$  as large as Brinkman and Rice's  $U_c$ .<sup>23), 24), 25)</sup> Since hysteresis appears, it is a discontinuous transition between the metallic and insulating phases at  $T = 0$  K. Since no jump occurs in the internal energy of Brinkman and Rice's theory, a plausible explanation for the hysteresis is that the residual entropy jumps at the transition. It is interesting to elucidate which is discontinuous at the transition point and is responsible for the hysteresis, the residual entropy, the internal energy, or some thing else.

Even if multisite terms are approximately or rigorously considered beyond  $S^3A$ , a problem of rigorously calculating all the single-site terms can also be substituted for another self-consistent Kondo problem,<sup>17), 18), 19), 26)</sup> in which the Anderson model to be solved has to be self-consistently determined with multisite terms as well as the single-site terms. Since the Anderson model is an effective Hamiltonian for the Kondo problem, a theory for treating either the self-consistent Kondo problem under or beyond  $S^3A$  is none other than a Kondo-lattice theory;<sup>17), 18), 19), 26)</sup> if multisite terms are considered beyond  $S^3A$ , the Kondo-lattice theory is none other than a  $1/D$  expansion theory. In the present paper, we list all the possible types of gapped phase of the Hubbard model under and beyond  $S^3A$  and then we examine on the basis of the Kondo-lattice or  $1/D$  expansion theory whether the residual entropy is zero or nonzero in each of them; we consider only possible phases within the constrained Hilbert subspace in which no symmetry is allowed to be broken, so that the possibility of the Mott insulator with symmetry broken or one with a hidden order is out of scope.

---

<sup>\*</sup>) In a single-band model such as the Hubbard model, three types of conventional Weiss mean field are of the zeroth order in  $1/D$ : spin density wave (SDW) or magnetism, charge density wave (CDW), and Bardeen-Cooper-Schrieffer (BCS) or isotopic  $s$ -wave superconductivity. In a multiband model, an orbital order in the charge or spin channel is also of the zeroth order in  $1/D$ .

The present paper is organized as follows: Preliminaries are given in §2. In §3, nongapped and gapped phases under  $S^3A$  are studied by the Kondo-lattice theory in order to confirm that, under  $S^3A$ , the Kondo-lattice theory is equivalent to DMFT. In §4, only possible complete-gapped phases beyond  $S^3A$  are studied by the Kondo-lattice theory. Discussion is given in §5. Conclusion is given in §6. Anomalies of possible gapped phases are studied in Appendix A. The entropy at  $T = +0$  K of possible complete-gapped phases is studied in Appendix B. Possible gapped phases in a nonlattice or continuous model, such as the Tomonaga-Luttinger model,<sup>(27), (28), (29)</sup> are studied in Appendix C.

## §2. Preliminaries

### 2.1. Model

We consider the Hubbard model on a hypercubic lattice in  $D$  dimensions:

$$\mathcal{H} = - \sum_{ij\sigma} t_{ij} d_{i\sigma}^\dagger d_{j\sigma} + U \sum_i n_{i\uparrow} n_{i\downarrow}, \quad (2.1)$$

where  $n_{i\sigma} = d_{i\sigma}^\dagger d_{i\sigma}$  and the other notations are conventional. The periodic boundary condition is assumed. When  $U = 0$ , the dispersion relation of an electron is given by

$$E(\mathbf{k}) = -\frac{1}{L} \sum_{ij} t_{ij} e^{i\mathbf{k} \cdot (\mathbf{x}_i - \mathbf{x}_j)}, \quad (2.2)$$

where  $\mathbf{k} = (k_1, k_2, \dots, k_D)$  is the wave number of the electron,  $\mathbf{x}_i = (i_1, i_2, \dots, i_D)a$ , with  $a$  being the lattice constant, is the position of the  $i$ th unit cell, and  $L$  is the number of unit cells. The hypercubic lattice is decomposed into two sublattices: one composed of unit cells with  $\sum_{\nu=1}^D i_\nu$  being even and the other composed of unit cells with  $\sum_{\nu=1}^D i_\nu$  being odd. For simplicity, we assume that the transfer integral  $t_{ij}$  is nonzero only between different sublattices, i.e., the Hubbard model is on a bipartite lattice. The density of states per unit cell is defined by

$$\rho_0(\varepsilon) = \frac{1}{L} \sum_{\mathbf{k}} \delta[\varepsilon - E(\mathbf{k})], \quad (2.3)$$

which is called the bare density of states in this paper. The band center of  $E(\mathbf{k})$  or  $\rho_0(\varepsilon)$  is denoted by  $\varepsilon_d$ ;  $\varepsilon_d = 0$  for the model of this paper. A hypersurface in the wave-number space,  $\mathbf{k}$  on which is denoted as  $\mathbf{k}_F$ , is defined by

$$E(\mathbf{k}_F) = \varepsilon_d = 0. \quad (2.4)$$

Since  $E(\mathbf{k})$  is symmetrical with respect to the hypersurface,  $\rho_0(\varepsilon)$  is symmetrical with respect to  $\varepsilon_d = 0$ :  $\rho_0(\varepsilon) = \rho_0(-\varepsilon)$ . We assume that the chemical potential  $\mu$  is exactly given by  $\mu = U/2$  and that the Hubbard model is exactly half-filled. In general, the particle-hole symmetry exists in a half-filled model on a bipartite lattice. The particle-hole symmetry is crucial in this study.

The transfer integral  $t_{ij}$  has to include the dimensional factor  $D$  in order for the effective bandwidth of  $E(\mathbf{k})$  to be finite for any  $D$ .<sup>(13),(14),(15),(16)</sup> For example, if  $t_{ij}$  is nonzero only between nearest neighbors,  $t_{ij} = t/\sqrt{D}$  between them. It follows that

$$E(\mathbf{k}) = -2\frac{t}{\sqrt{D}} \sum_{\nu=1}^D \cos(k_{\nu}a). \quad (2.5)$$

The absolute bandwidth of  $E(\mathbf{k})$  is given by  $W = 4|t|\sqrt{D}$ . Since  $(1/L) \sum_{\mathbf{k}} E(\mathbf{k}) = 0$  and  $(1/L) \sum_{\mathbf{k}} E^2(\mathbf{k}) = 2t^2$ , the effective bandwidth of  $E(\mathbf{k})$  is  $O(|t|)$  for any  $D$ . If  $t_{ij}$  is of finite range, either the absolute or effective bandwidth is finite for any finite  $D$ . We call any model of this type for a finite  $D$  a finite bandwidth model.

If  $t_{ij}$  is of infinitely long range, it is possible that  $E(\mathbf{k})$  is given by

$$E(\mathbf{k}) = \sum_{\nu=1}^D \frac{|t_{\infty}|}{D} \eta(k_{\nu}a), \quad (2.6)$$

where

$$\eta(x) = \begin{cases} \tan(x - \pi/2), & 0 < x < \pi, \\ -\tan(x + \pi/2), & -\pi < x < 0. \end{cases} \quad (2.7)$$

The bare density of states is of a Lorentzian type for any  $D$ :

$$\rho_0(\varepsilon) = \frac{1}{L} \sum_{\mathbf{k}} \delta[\varepsilon - E(\mathbf{k})] = \frac{1}{\pi} \frac{|t_{\infty}|}{\varepsilon^2 + |t_{\infty}|^2}. \quad (2.8)$$

We call this model an infinite bandwidth model.

## 2.2. Kondo-lattice theory

We constrain the Hilbert space within the subspace where no symmetry is allowed to be broken. We first assume that  $L$  is large but still finite and that the temperature is nonzero such that  $T > 0$  K; the thermodynamic limit of  $L \rightarrow +\infty$ , if necessary, followed by the low-temperature limit of  $T \rightarrow 0$  K is taken last.

The site representation is convenient for formulating the Kondo-lattice theory. The Green function and self-energy in the grand canonical ensemble are denoted by  $R_{ij;\sigma}(i\varepsilon_l)$  and  $\Sigma_{ij;\sigma}(i\varepsilon_l)$ , respectively, where  $\varepsilon_l = (2l+1)\pi k_B T$ , with  $l$  being an integer, is a fermionic energy. We consider a skeleton Feynman diagram for  $\Sigma_{ij;\sigma}(i\varepsilon_l)$ , but with each electron line replaced by a renormalized line of  $R_{ij;\sigma}(i\varepsilon_l)$ . If only lines of  $U$  and site-diagonal  $R_{ii\sigma}(i\varepsilon_l)$  appear in it, it is a single-site diagram; if at least one line of site-off-diagonal  $R_{i \neq j;\sigma}(i\varepsilon_l)$  appears in it, it is a multisite diagram. All the Feynman diagrams can be classified into the single-site and multisite ones without any ambiguity. Thus,

$$\Sigma_{ij;\sigma}(i\varepsilon_l) = \Sigma_{i;\sigma}(i\varepsilon_l)\delta_{ij} + \Delta\Sigma_{ij;\sigma}(i\varepsilon_l), \quad (2.9)$$

where  $\Sigma_{i;\sigma}(i\varepsilon_l)$  is the single-site self-energy,  $\delta_{ij}$  is the Kronecker delta, and  $\Delta\Sigma_{ij;\sigma}(i\varepsilon_l)$  is the multisite self-energy. Since the site-diagonal  $R_{ii;\sigma}(i\varepsilon_l)$  and the single-site

$\Sigma_{i;\sigma}(i\varepsilon_l)$  do not depend on the site index  $i$ , they are simply denoted by  $R_\sigma(i\varepsilon_l)$  and  $\Sigma_\sigma(i\varepsilon_l)$ , respectively, in the following part of this paper:

The single-site  $\Sigma_\sigma(i\varepsilon_l)$  is of the zeroth or leading order in  $1/D$ , while the multisite  $\Delta\Sigma_{ij;\sigma}(i\varepsilon_l)$  is of higher order in  $1/D$ .<sup>(13),(14),(15),(16)</sup> Any type of conventional Weiss mean field is a multisite effect. If no multisite term is considered, the Kondo-lattice theory is reduced to  $S^3A$ , which is also DMFT or DCPA. In this paper, the multisite  $\Delta\Sigma_{ij;\sigma}(i\varepsilon_l)$  is considered, but no conventional Weiss mean field is considered.

If the multisite  $\Delta\Sigma_{ij;\sigma}(i\varepsilon_l)$  is ignored under  $S^3A$  or if it is approximately or rigorously considered beyond  $S^3A$ , the single-site  $\Sigma_\sigma(i\varepsilon_l)$  is exactly equal to the local self-energy  $\tilde{\Sigma}_\sigma(i\varepsilon_l)$  of the Anderson model that is determined by solving the self-consistent Kondo problem, as discussed below. The mapping of  $\Sigma_\sigma(i\varepsilon_l)$  to  $\tilde{\Sigma}_\sigma(i\varepsilon_l)$ , such as

$$\Sigma_\sigma(i\varepsilon_l) = \tilde{\Sigma}_\sigma(i\varepsilon_l), \quad (2.10)$$

is the key of the Kondo-lattice theory. The Anderson model is defined by

$$\tilde{\mathcal{H}} = \tilde{\epsilon}_d \sum_{\sigma} \tilde{n}_{d\sigma} + \sum_{\mathbf{k}\sigma} \tilde{E}_c(\mathbf{k}) \tilde{c}_{\mathbf{k}\sigma}^\dagger \tilde{c}_{\mathbf{k}\sigma} + \frac{1}{\sqrt{\tilde{L}}} \sum_{\mathbf{k}\sigma} \left( \tilde{V}_{\mathbf{k}} \tilde{c}_{\mathbf{k}\sigma}^\dagger \tilde{d}_\sigma + \tilde{V}_{\mathbf{k}}^* \tilde{d}_\sigma^\dagger \tilde{c}_{\mathbf{k}\sigma} \right) + \tilde{U} \tilde{n}_{d\uparrow} \tilde{n}_{d\downarrow}, \quad (2.11)$$

where  $\tilde{n}_{d\sigma} = \tilde{d}_\sigma^\dagger \tilde{d}_\sigma$  and the notations are conventional except for the use of tildes. The local Green function for localized electrons in the Anderson model is given by

$$\tilde{G}_\sigma(i\varepsilon_l) = \frac{1}{i\varepsilon_l + \tilde{\mu} - \tilde{\epsilon}_d - \tilde{\Sigma}_\sigma(i\varepsilon_l) - \tilde{\Gamma}(i\varepsilon_l)}, \quad (2.12)$$

where  $\tilde{\mu}$  is the chemical potential,  $\tilde{\Sigma}_\sigma(i\varepsilon_l)$  is the local self-energy, and

$$\tilde{\Gamma}(i\varepsilon_l) = \frac{1}{\pi} \int_{-\infty}^{+\infty} d\epsilon \frac{\tilde{\Delta}(\epsilon)}{i\varepsilon_l - \epsilon}, \quad (2.13a)$$

with

$$\tilde{\Delta}(\epsilon) = \frac{\pi}{\tilde{L}} \sum_{\mathbf{k}} |\tilde{V}_{\mathbf{k}}|^2 \delta[\epsilon + \tilde{\mu} - \tilde{E}_c(\mathbf{k})], \quad (2.13b)$$

is the hybridization term with conduction electrons; and  $\tilde{\Delta}(\epsilon) = -\text{Im}\tilde{\Gamma}(\epsilon + i0)$ . Although the Anderson model is characterized by  $\tilde{\mu} - \tilde{\epsilon}_d$ ,  $\tilde{U}$ ,  $\tilde{E}_c(\mathbf{k})$ , and  $\tilde{V}_{\mathbf{k}}$ , it is essentially uniquely characterized only by  $\tilde{\mu} - \tilde{\epsilon}_d$ ,  $\tilde{U}$ , and  $\tilde{\Delta}(\epsilon)$ . In the Kondo-lattice theory, not only the local  $\tilde{\Sigma}_\sigma(i\varepsilon_l)$  of the Anderson model and the multisite  $\Delta\Sigma_{ij;\sigma}(i\varepsilon_l)$  of the Hubbard model, but also the  $\tilde{\mu} - \tilde{\epsilon}_d$ ,  $\tilde{U}$ , and  $\tilde{\Delta}(\epsilon)$  of the Anderson model have to be self-consistently calculated or determined to satisfy Eq. (2.10); the Anderson model determined in this way is called the mapped Anderson model.

Even if any arbitrary diagram for  $\Sigma_\sigma(i\varepsilon_l)$  of the Hubbard model is considered, the very same one for  $\tilde{\Sigma}_\sigma(i\varepsilon_l)$  of the Anderson model is obtained. This one-to-one correspondence between the diagrams for  $\Sigma_\sigma(i\varepsilon_l)$  and  $\tilde{\Sigma}_\sigma(i\varepsilon_l)$  is unique, and there is no lonely diagram in either  $\Sigma_\sigma(i\varepsilon_l)$  or  $\tilde{\Sigma}_\sigma(i\varepsilon_l)$ . In each pair of diagrams,

the interaction and electron lines are  $U$  and  $R_\sigma(i\varepsilon_l)$ , respectively, for  $\Sigma_\sigma(i\varepsilon_l)$  of the Hubbard model, while they are  $\tilde{U}$  and  $\tilde{G}_\sigma(i\varepsilon_l)$ , respectively, for  $\tilde{\Sigma}_\sigma(i\varepsilon_l)$  of the Anderson model. Provided that a couple of

$$\tilde{U} = U, \quad (2.14a)$$

and

$$\tilde{G}_\sigma(i\varepsilon_l) = R_\sigma(i\varepsilon_l), \quad (2.14b)$$

are satisfied, the contributions of any pair of diagrams are exactly the same as each other; thus, the mapping of Eq. (2.10) is satisfied. The condition of Eq. (2.14b) is equivalent to a couple of

$$\tilde{\mu} - \tilde{\epsilon}_d = \mu - \epsilon_d, \quad (2.14c)$$

where  $\epsilon_d$  is the band center and  $\epsilon_d = 0$ , and

$$\tilde{\Delta}(\varepsilon) = \text{Im} \left[ \tilde{\Sigma}_\sigma(\varepsilon + i0) + 1/R_\sigma(\varepsilon + i0) \right], \quad (2.14d)$$

where  $\tilde{\Sigma}_\sigma(\varepsilon + i0)$  and  $R_\sigma(\varepsilon + i0)$  are the analytical continuations of  $\tilde{\Sigma}_\sigma(i\varepsilon_l)$  and  $R_\sigma(i\varepsilon_l)$ , respectively, from the upper-half complex plane onto the real axis. Since Eqs. (2.14a) and (2.14c) are trivial, Eq. (2.14d) is a practical mapping condition for determining the mapped Anderson model.\*)

If a series of Feynman diagrams, single-site or multisite ones, for the Hubbard model are divergent in the thermodynamic limit of  $L \rightarrow +\infty$  followed by  $T \rightarrow 0$  K, the limit has to be taken at this stage. The mapped Anderson model depends on  $L$  and  $T$ . In the following part of this paper, we consider the mapped Anderson model in the limit of  $L \rightarrow +\infty$ , if necessary, followed by  $T \rightarrow 0$  K.

In the wave-number representation, the self-energy is given by

$$\Sigma_\sigma(i\varepsilon_l, \mathbf{k}) = \tilde{\Sigma}_\sigma(i\varepsilon_l) + \Delta\Sigma_\sigma(i\varepsilon_l, \mathbf{k}), \quad (2.15)$$

where Eq. (2.10) is used for the single-site self-energy, and

$$\Delta\Sigma_\sigma(i\varepsilon_l, \mathbf{k}) = \frac{1}{L} \sum_{ij} e^{-i\mathbf{k} \cdot (\mathbf{x}_i - \mathbf{x}_j)} \Delta\Sigma_{ij;\sigma}(i\varepsilon_l). \quad (2.16)$$

The Green function is given by

$$G_\sigma(i\varepsilon_l, \mathbf{k}) = \frac{1}{L} \sum_{ij} e^{-i\mathbf{k} \cdot (\mathbf{x}_i - \mathbf{x}_j)} R_{ij;\sigma}(i\varepsilon_l) = \frac{1}{i\varepsilon_l + \mu - E(\mathbf{k}) - \Sigma_\sigma(i\varepsilon_l, \mathbf{k})}. \quad (2.17)$$

The site-diagonal Green function is given by

$$R_\sigma(i\varepsilon_l) = \frac{1}{L} \sum_{\mathbf{k}} \frac{1}{i\varepsilon_l + \mu - E(\mathbf{k}) - \Sigma_\sigma(i\varepsilon_l, \mathbf{k})}. \quad (2.18)$$

---

\*) Following the previous paper,<sup>26)</sup> it can be proved on the basis of Eq. (2.14d) that  $\tilde{\Delta}(\varepsilon) \geq 0$  for any possible self-consistent solution, as anticipated.

The density of states is simply given by

$$\rho(\varepsilon) = -\frac{1}{\pi} \text{Im} R_\sigma(\varepsilon + i0). \quad (2.19a)$$

Because of Eq. (2.14b), it is equal to that of the mapped Anderson model either under or beyond S<sup>3</sup>A:

$$\rho(\varepsilon) = -\frac{1}{\pi} \text{Im} R_\sigma(\varepsilon + i0) = -\frac{1}{\pi} \text{Im} \tilde{G}_\sigma(\varepsilon + i0). \quad (2.19b)$$

This equality is equivalent to the mapping condition and is crucial in this paper.

When  $\mu = U/2$ , not only the Hubbard model but also the mapped Anderson model is half-filled and symmetrical:

$$\mu = \tilde{\mu} - \tilde{\epsilon}_d = U/2 = \tilde{U}/2, \quad \langle \mathcal{N} \rangle / L = \langle \tilde{n}_\uparrow + \tilde{n}_\downarrow \rangle = 1, \quad (2.20)$$

where  $\mathcal{N} = \sum_{i\sigma} n_{i\sigma}$  and  $\langle \cdots \rangle$  stands for the statistical average. Since the particle-hole symmetry exists in any half-filled symmetrical model, various equalities are inevitably satisfied: In particular,

$$\text{Re} \tilde{\Sigma}_\sigma(+i0) - \mu = 0, \quad \text{Re} \tilde{\Gamma}(+i0) = 0, \quad E(\mathbf{k}_F) + \text{Re} \Sigma_\sigma(+i0, \mathbf{k}_F) - \mu = 0, \quad (2.21)$$

are crucial, where  $\mathbf{k}_F$  is defined by Eq. (2.4) or  $E(\mathbf{k}_F) = 0$ .

It is easy to see from Eqs. (2.12), (2.18), (2.19), and (2.21) that, unless  $\text{Im} [\tilde{\Sigma}_\sigma(\varepsilon + i0) + \tilde{\Gamma}(\varepsilon + i0)]$  is divergent at  $\varepsilon = 0$ , then  $\rho(0) > 0$  and not  $\rho(0) = 0$  in the Anderson model and that, unless  $\text{Im} \Sigma_\sigma(\varepsilon + i0, \mathbf{k}_F)$  is divergent at  $\varepsilon = 0$ , then  $\rho(0) > 0$  and not  $\rho(0) = 0$  in the Hubbard model. A couple of these features are also crucial in this paper.

### 2.3. Kondo effect

#### 2.3.1. Normal Fermi liquid in the Anderson model

The Kondo problem has already been solved;<sup>(30),(31),(32),(33),(34),(35),(36)</sup> in particular, the Bethe-ansatz solution was given to the Anderson model in which  $\tilde{\Delta}(\varepsilon)$  is constant as well as to the  $s$ - $d$  model in which the density of states of conduction electrons is constant.<sup>(37),(38),(39),(40)</sup> For example, we consider the Anderson model defined by Eq. (2.11). The static spin susceptibility of localized electrons is given by

$$\tilde{\chi}_s(T) = (1/2)g^2\mu_B^2 [\tilde{\chi}_{\uparrow\uparrow}(T) - \tilde{\chi}_{\uparrow\downarrow}(T)], \quad (2.22)$$

where  $g$  and  $\mu_B$  are the  $g$  factor and Bohr magneton, respectively, of localized electrons, and

$$\tilde{\chi}_{\sigma\sigma'}(T) = \int_0^{1/k_B T} d\tau \left\langle e^{\tau(\tilde{\mathcal{H}} - \tilde{\mu}\tilde{\mathcal{N}})} \tilde{n}_{d\sigma} e^{-\tau(\tilde{\mathcal{H}} - \tilde{\mu}\tilde{\mathcal{N}})} \tilde{n}_{d\sigma'} \right\rangle, \quad (2.23)$$

where  $\tilde{\mathcal{N}} = \sum_\sigma \tilde{n}_{d\sigma} + \sum_{\mathbf{k}\sigma} \tilde{c}_{\mathbf{k}\sigma}^\dagger \tilde{c}_{\mathbf{k}\sigma}$ . Since the most crucial physics of the Kondo effect is the quenching of a local moment by conduction electrons,<sup>(30),(31),(32)</sup> determining whether the Fermi surface of conduction electrons exists is crucial. The Fermi surface

is defined by  $\tilde{\mu} = \tilde{E}_c(\mathbf{k})$ . If the Fermi surface exists and  $|\tilde{V}_{\mathbf{k}}|^2$  is nonzero at least on a part of the Fermi surface, it follows that

$$0 < \tilde{\Delta}(\varepsilon) < +\infty, \quad (2.24)$$

at least for  $\varepsilon \simeq 0$ , so that the ground state is a singlet and a normal Fermi liquid. The Kondo temperature  $T_K$  is defined by<sup>\*)</sup>

$$k_B T_K = (1/4)g^2 \mu_B^2 / \tilde{\chi}_s(0). \quad (2.25)$$

If  $T \ll T_K$ ,  $\tilde{\chi}_s(T)$  exhibits the Pauli paramagnetism; if  $T \gg T_K$ ,  $\tilde{\chi}_s(T)$  obeys the Curie-Weiss law, i.e., localized electrons behave like a local moment or a free spin.

If  $\tilde{\Delta}(0)$  is neither zero nor infinite and if  $T_K$  is nonzero, the ground state is a normal Fermi liquid, so that the self-energy of localized electrons is analytical at  $\varepsilon = 0$  and it can be expanded in a way such that<sup>34), 35), 36)</sup>

$$\tilde{\Sigma}_\sigma(\varepsilon + i0) = \frac{1}{2}U + (1 - \tilde{\phi}_1)\varepsilon - i\tilde{\phi}_2\varepsilon^2 / \tilde{\Delta}(0) - i\tilde{\phi}_3 (k_B T)^2 / \tilde{\Delta}(0) + \dots, \quad (2.26)$$

for  $|\varepsilon| \ll k_B T_K$  and  $T \ll T_K$ , where  $\tilde{\phi}_1 > 1$ ,  $\tilde{\phi}_2 > 0$ , and  $\tilde{\phi}_3 > 0$ . When  $T = 0$  K, according to Eqs. (2.12), (2.19), (2.21), and (2.26), the density of states of localized electrons at the chemical potential is simply given by

$$\rho(0) = 1 / [\pi \tilde{\Delta}(0)]. \quad (2.27)$$

If  $\tilde{\Delta}(\varepsilon)$  is constant such that  $\tilde{\Delta}(\varepsilon) = \tilde{\Delta}(0)$ , Anderson's compensation theorem is valid;<sup>41)</sup> there is no additional contribution of conduction electrons when either  $\tilde{V}_{\mathbf{k}}$  or  $\tilde{U}$  is introduced. Then, the expansion coefficients are simply given by<sup>34), 35), 36)</sup>

$$\tilde{\phi}_1 = \tilde{\chi}_{\uparrow\uparrow}(0) [\pi \tilde{\Delta}(0)], \quad \tilde{\phi}_2 = \tilde{\phi}_3 = \left\{ \tilde{\chi}_{\uparrow\downarrow}(0) [\pi \tilde{\Delta}(0)] \right\}^2 / 2. \quad (2.28)$$

The specific-heat coefficient of localized electrons is given by

$$\tilde{\gamma} = (2/3)\pi^2 k_B^2 \tilde{\phi}_1 / [\pi \tilde{\Delta}(0)]. \quad (2.29)$$

The Fermi-liquid relation above is consistent with the Bethe-ansatz solution.<sup>37), 38), 39), 40)</sup>

If  $\tilde{\Delta}(\varepsilon)$  depends on  $\varepsilon$ , the compensation theorem is not valid, in general; there are additional contributions of conduction electrons. If the compensation theorem is valid or not valid, the static spin susceptibility and specific-heat coefficient of localized electrons, which include no additional contribution of conduction electrons, are given by Eqs. (2.22) and (2.29), respectively, but with  $\tilde{\phi}_1 \neq \tilde{\chi}_{\uparrow\uparrow}(0) [\pi \tilde{\Delta}(0)]$  in general. The Wilson ratio is defined by

$$\tilde{W}_s = \frac{\tilde{\chi}_s(0)}{\tilde{\gamma}} \frac{(2/3)\pi^2 k_B^2}{(1/2)g^2 \mu_B^2}, \quad (2.30)$$

which is of the order of unity or  $O(1)$ , at least, when the onsite  $\tilde{U}$  is repulsive.

---

<sup>\*)</sup> This definition of  $T_K$  is different from Wilson's<sup>32)</sup> by a numerical factor, even in the  $s$ - $d$  limit.



If  $T_K$  is nonzero, the ground state of the Anderson model is a normal Fermi liquid. When  $T = 0$  K,  $\text{Im}\tilde{\Sigma}_\sigma(+i0) = 0$  and the entropy or residual entropy of localized electrons is zero. When  $0 \text{ K} < T \ll T_K$ ,  $\text{Im}\tilde{\Sigma}_\sigma(+i0)$  is negatively nonzero and small, and the entropy of localized electrons is much smaller than  $k_B \ln 2$ . When  $T \ll T_K$ , localized electrons never behave like a free spin. When  $T \gtrsim T_K$ ,  $\text{Im}\tilde{\Sigma}_\sigma(+i0)$  is negatively large, the entropy of localized electrons is as large as  $k_B \ln 2$ , and localized electrons behave like a free spin. This crossover as a function of  $T$  is crucial.<sup>32)</sup>

### 2.3.2. Possible gapped phases in the Anderson model

If Eq. (2.24) is not satisfied, it is possible that  $T_K = 0$  K or  $T_K = +0$  K. If  $T_K = 0$  K or  $T_K = +0$  K, the self-energy  $\tilde{\Sigma}_\sigma(\varepsilon + i0)$  is singular at  $\varepsilon = 0$ . We examine how singular  $\tilde{\Sigma}_\sigma(\varepsilon + i0)$  is in the case of  $T_K = 0$  K or  $T_K = +0$  K.

According to Hubbard's theory,<sup>6)</sup> in the so-called atomic limit of  $W/U \rightarrow 0$ , the self-energy of the Hubbard model is given by

$$\Sigma_\sigma(\varepsilon + i0) = \frac{1}{2}U + \frac{U^2}{4} \frac{1}{\varepsilon + i0}. \quad (2.31)$$

Since Hubbard's theory is under SSA, it is anticipated on the basis of S<sup>3</sup>A or the Kondo-lattice theory that  $\Sigma_\sigma(\varepsilon + i0)$  given by Eq. (2.31) can be approximately used for  $\tilde{\Sigma}_\sigma(\varepsilon + i0)$  of the Anderson model. The atomic limit in the Hubbard model corresponds to the limit of  $T_K \rightarrow 0$  K or the case of  $T_K = 0$  K in the Anderson model. It is anticipated that, if  $T_K = +0$  K or  $T_K = 0$  K, the self-energy  $\tilde{\Sigma}_\sigma(\varepsilon + i0)$  of the Anderson model has a pole at  $\varepsilon = 0$ .

Next, we examine how the self-energy  $\tilde{\Sigma}_\sigma(\varepsilon + i0)$  of the Anderson model becomes singular as  $T_K \rightarrow 0$  K. For example, we consider a function defined by

$$\tilde{\Sigma}_\sigma(\varepsilon + i0; \gamma_0) = \frac{1}{2}U + \frac{U^2}{4} \int_{+0}^{+\infty} \frac{d\gamma}{\gamma_0} \int_{+0}^{\pi/2} d\theta \zeta(\gamma/\gamma_0, \theta) \Xi(\varepsilon; \gamma, \theta), \quad (2.32)$$

where  $\gamma_0 \simeq k_B T_K$  is a parameter,  $\zeta(x, \theta)$  is a functional parameter, and

$$\Xi(\varepsilon; \gamma, \theta) = \frac{1}{2 \cos \theta} \left[ \frac{-e^{-i(\pi-\theta)}}{\varepsilon - \gamma e^{-i(\pi-\theta)}} + \frac{-e^{-i\theta}}{\varepsilon - \gamma e^{-i\theta}} \right] = \frac{-\varepsilon(\gamma^2 - \varepsilon^2) - i2\varepsilon^2 \gamma \sin \theta}{\varepsilon^4 - 2\varepsilon^2 \gamma^2 \cos 2\theta + \gamma^4}. \quad (2.33)$$

Here, we assume that  $\zeta(x, \theta)$  satisfies  $\zeta(x, \theta) \geq 0$ ,  $\zeta(x \rightarrow +\infty, \theta) = 0$ , and

$$\int_{+0}^{+\infty} dx \int_{+0}^{\pi/2} d\theta \zeta(x, \theta) = \int_{+0}^{+\infty} dx \int_{+0}^{\pi/2} d\theta \zeta(x, \theta) x = 1. \quad (2.34)$$

It is easy to see that  $\Xi(\varepsilon; \gamma, \theta)$  as a function of  $\varepsilon$  is analytical in the upper-half complex plane at least for  $\gamma > 0$  and  $0 < \theta < \pi/2$ . It should be noted that

$$\frac{1}{\pi} \int_{-\infty}^{+\infty} d\varepsilon \frac{2\gamma\varepsilon^2 \sin \theta}{\varepsilon^4 - 2\varepsilon^2 \gamma^2 \cos 2\theta + \gamma^4} = 1. \quad (2.35)$$

Thus,  $-(1/\pi)\text{Im}\Xi(\varepsilon; \gamma \rightarrow 0, \theta)$  is the delta function  $\delta(\varepsilon)$  or, at least, it can be treated as the delta function, although  $\text{Im}\Xi(\varepsilon = 0; \gamma, \theta) = 0$  for any nonzero  $\gamma$ . It follows that

$$\lim_{\varepsilon/\gamma_0 \rightarrow 0} \tilde{\Sigma}_\sigma(\varepsilon + i0; \gamma_0) = \frac{1}{2}U - \frac{U^2}{4\gamma_0^2} \left( \tilde{s}_1\varepsilon + i\tilde{s}_2\frac{\varepsilon^2}{\gamma_0} + \cdots \right), \quad (2.36a)$$

$$\lim_{\varepsilon/\gamma_0 \rightarrow \pm\infty} \tilde{\Sigma}_\sigma(\varepsilon + i0; \gamma_0) = \frac{1}{2}U + \frac{U^2}{4} \left( \frac{1}{\varepsilon} - i\tilde{s}_3\frac{\gamma_0}{\varepsilon^2} + \cdots \right), \quad (2.36b)$$

where

$$\tilde{s}_1 = \int_{+0}^{+\infty} dx \int_{+0}^{\pi/2} d\theta \zeta(x, \theta) \frac{1}{x^2}, \quad \tilde{s}_2 = \int_{+0}^{+\infty} dx \int_{+0}^{\pi/2} d\theta \zeta(x, \theta) \frac{2 \sin \theta}{x^3}, \quad (2.37a)$$

$$\tilde{s}_3 = \int_{+0}^{+\infty} dx \int_{+0}^{\pi/2} d\theta \zeta(x, \theta) 2x \sin \theta. \quad (2.37b)$$

If  $\gamma_0 \simeq k_B T_K$  and  $\zeta(x, \theta)$  are properly chosen,  $\tilde{\Sigma}_\sigma(\varepsilon + i0; \gamma_0)$  behaves in almost the same way as  $\tilde{\Sigma}_\sigma(\varepsilon + i0)$  given by Eq. (2.26) for  $|\varepsilon| \ll \gamma_0$  or  $|\varepsilon| \ll k_B T_K$ . On the other hand, it follows from Eqs. (2.35) and (2.36b) that

$$\lim_{\gamma_0 \rightarrow 0} \lim_{\varepsilon/\gamma_0 \rightarrow \pm\infty} \tilde{\Sigma}_\sigma(\varepsilon + i0; \gamma_0) = \frac{1}{2}U + \frac{U^2}{4} \frac{1}{\varepsilon + i0}. \quad (2.38)$$

Thus,  $\tilde{\Sigma}_\sigma(\varepsilon + i0; \gamma_0 \rightarrow 0)$  has a pole or a variation of the pole at  $\varepsilon = 0$ ; the pole is not simple if  $\gamma_0$  is nonzero. It is anticipated that the analytical self-energy becomes singular in this way or a similar way as  $\gamma_0 \rightarrow 0$  or  $T_K \rightarrow 0$  K.

It is anticipated on the basis of the arguments above that there is no essential difference between the self-energy characterized by  $T_K = 0$  K and that by  $T_K = +0$  K. If the self-energy  $\tilde{\Sigma}_\sigma(\varepsilon + i0)$  has a pole or the variation of the pole discussed above at  $\varepsilon = 0$ , if the residue of the pole is large enough, if  $\text{Im}\tilde{\Sigma}_\sigma(\varepsilon + i0) = 0$  for  $|\varepsilon| < \epsilon_g/2$  but  $\varepsilon = 0$ , and if  $\tilde{\Delta}(\varepsilon) = 0$  for  $|\varepsilon| < \epsilon_g/2$ , a gap as large as  $\epsilon_g$  opens in the Anderson model.<sup>\*)</sup>

To be precise, it has never been fully clarified what singularity the self-energy  $\tilde{\Sigma}_\sigma(\varepsilon + i0)$  of the Anderson model has to have at  $\varepsilon = 0$  in the case of  $T_K = 0$  K and  $T_K = +0$  K, i.e., a simple pole, a variation of the pole, or some thing else. However, at least, it is certain that, unless  $\tilde{\Delta}(0)$  is exactly zero or infinite and unless  $T_K$  is the absolute zero or infinitesimal, the self-energy  $\tilde{\Sigma}_\sigma(\varepsilon + i0)$  has no type of singularity at  $\varepsilon = 0$  and it can be expanded as in Eq. (2.26).

Unless  $\tilde{\Delta}(0)$  is infinite,<sup>\*\*)</sup> a gap can open only if at least  $\text{Im}\tilde{\Sigma}_\sigma(\varepsilon + i0)$  is divergent at  $\varepsilon = 0$ . The divergence of  $\text{Im}\tilde{\Sigma}_\sigma(\varepsilon + i0)$  at  $\varepsilon = 0$  is possible only if  $T_K = 0$  K or  $T_K = +0$  K. If  $T_K = 0$  K, the residual entropy is nonzero; localized electrons behave like a free spin even at  $T = 0$  K. If  $T = +0$  K, the entropy is zero at  $T = 0$  K but is

---

<sup>\*)</sup> According to Hubbard's theory,<sup>6), 7)</sup> when  $U/W$  is so large that a Hubbard gap as large as  $\epsilon_g$  opens,  $\Sigma_\sigma(\varepsilon + i0)$  has a pole at  $\varepsilon = 0$ , the residue of the pole is large enough, and  $\text{Im}\Sigma_\sigma(\varepsilon + i0) = 0$  for  $|\varepsilon| < \epsilon_g/2$  but  $\varepsilon = 0$ , as in the Anderson model.

<sup>\*\*)</sup> If  $\tilde{\Delta}(0)$  is divergent, a gap inevitably opens.

nonzero at  $T = +0\text{ K}$ ; localized electrons behave like a free spin even at  $T = +0\text{ K}$ . The third law of thermodynamics is broken in any gapped phase of the Anderson model characterized by  $T_K = 0\text{ K}$  or  $T_K = +0\text{ K}$ .<sup>\*)</sup>

### 2.3.3. Kondo temperature of the mapped Anderson model

According to the Kondo-lattice theory, the Kondo temperature  $T_K$  or  $k_B T_K$  of the mapped Anderson model is also the energy scale of single-site quantum spin fluctuations in the Hubbard model.<sup>\*\*)</sup>  Thus, we also call the energy scale in the Hubbard model the Kondo temperature; it is also denoted by  $T_K$  or  $k_B T_K$ . If the Kondo temperature  $T_K$  of the mapped Anderson model is the absolute zero or infinitesimal or if the self-energy  $\tilde{\Sigma}_\sigma(\varepsilon + i0)$  of the mapped Anderson model is divergent at  $\varepsilon = 0$ , it can be concluded that at least the single-site part of the residual entropy or entropy at  $T = +0\text{ K}$  is nonzero in the Hubbard model.

## §3. Nongapped and gapped phases under $S^3A$

### 3.1. Infinite bandwidth model

We consider under  $S^3A$  the infinite bandwidth model in  $D \geq 1$  dimensions, in which the bare density of states  $\rho_0(E)$  is given by Eq. (2.8). Since  $\rho_0(E)$  is of a Lorentzian type, the problem is exactly the same as that for the Bethe-lattice model in  $D \rightarrow +\infty$  dimensions,<sup>20)</sup> which is exactly solvable.

The multisite  $\Delta\Sigma_\sigma(i\varepsilon_l, \mathbf{k})$  vanishes under  $S^3A$ . It follows from Eq. (2.18) that

$$R_\sigma(\varepsilon + i0) = \int_{-\infty}^{+\infty} dE \rho_0(E) \frac{1}{\varepsilon + \mu - E - \tilde{\Sigma}_\sigma(\varepsilon + i0)} \quad (3.1a)$$

$$= \frac{1}{\varepsilon + \mu - \tilde{\Sigma}_\sigma(\varepsilon + i0) + i|t_\infty|}. \quad (3.1b)$$

It follows from Eq. (2.14d) that  $\tilde{\Delta}(\varepsilon) = |t_\infty|$ . The self-consistent Kondo problem is reduced to a *non-self-consistent* Kondo problem, to which the Bethe-ansatz solution was given.<sup>37),38),39),40)</sup> The ground state is a normal Fermi liquid;  $T_K$  is nonzero. The density of states is simply given by  $\rho(\varepsilon) = -(1/\pi)\text{Im}R_\sigma(\varepsilon + i0)$ , where  $R_\sigma(\varepsilon + i0)$  is given by Eq. (3.1b). Since the single-site  $\tilde{\Sigma}_\sigma(\varepsilon + i0)$  can be expanded as shown in Eq. (2.26), the density of states in the vicinity of the chemical potential is approximately given by that of a Lorentzian type, whose bandwidth is  $O(k_B T_K)$ . In particular,  $\rho(0) = \rho_0(0) = 1/(\pi|t_\infty|)$  at  $T = 0\text{ K}$ . When  $T = 0\text{ K}$ ,  $\rho(0)$  is never renormalized by  $U$ .

---

<sup>\*)</sup> One of the expressions of the third law of thermodynamics is that the residual entropy has to be zero. Another expression is that the absolute zero Kelvin or  $T = 0\text{ K}$  can never be reached. Since  $T = 0\text{ K}$  can never be reached, it is possible to extend the first expression to the third one: Entropy has to be continuously decreasing and vanishing as  $T \rightarrow 0\text{ K}$ . If the possibility of an ordered ground state in an actual system in three dimensions is considered, the third expression of the third law can be extended to the fourth one: Entropy has to be continuously decreasing and vanishing faster than linearly in  $T$  as  $T \rightarrow 0\text{ K}$ .

<sup>\*\*)</sup>  Since the mapped Anderson model is determined for a given  $T$ , it depends on the given  $T$ , in general. Thus, the Kondo temperature  $T_K$  in the Hubbard model depends on  $T$ , in general.

### 3.2. Finite bandwidth model

When the bare bandwidth  $W$  is finite and the onsite  $U$  is large enough such that  $U \gtrsim W$ , a complete gap opens at  $T = 0$  K under DMFT,<sup>(23),(24),(25)</sup> which is simply  $S^3A$ . In this subsection, we list all the possible types of solution under  $S^3A$  and we examine the nature of each of them. We do not specify a precise model for  $\rho_0(E)$ ; we only assume that  $\rho_0(E) > 0$  for  $|E| < W/2$  and  $\rho_0(E) = 0$  for  $|E| > W/2$ .

The density of states is given by Eq. (3.1a) for the Hubbard model and

$$\rho(\varepsilon) = -\frac{1}{\pi} \text{Im} \frac{1}{\varepsilon + \mu - \tilde{\Sigma}_\sigma(\varepsilon + i0) - \tilde{I}(\varepsilon + i0)}, \quad (3.2)$$

for the mapped Anderson model. Under  $S^3A$ ,  $\tilde{\Sigma}_\sigma(\varepsilon + i0)$  and  $\tilde{I}(\varepsilon + i0)$  are self-consistently calculated or determined in a way such that Eqs. (3.1a) and (3.2) are equal to each other. According to Eq. (2.13), if one of the real and imaginary parts of  $\tilde{I}(\varepsilon + i0)$  is finite and continuous at  $\varepsilon = \varepsilon_0$ , the other is also finite and continuous at  $\varepsilon = \varepsilon_0$ ; and if one of them is discontinuous or divergent at  $\varepsilon = \varepsilon_0$ , the other is divergent or discontinuous at  $\varepsilon = \varepsilon_0$ . In general,

$$\tilde{\Sigma}_\sigma(\varepsilon + i0) - \mu = -\frac{1}{\pi} \int_{-\infty}^{+\infty} dx \frac{\text{Im} \tilde{\Sigma}_\sigma(x + i0)}{\varepsilon + i0 - x}, \quad (3.3)$$

for any symmetrical model. According to Eq. (3.3), the real and imaginary parts of  $\tilde{\Sigma}_\sigma(\varepsilon + i0)$  satisfy the very same relation as that for  $\tilde{I}(\varepsilon + i0)$  discussed above.

First, we consider a possible solution in which  $\text{Re} \tilde{\Sigma}_\sigma(\varepsilon + i0)$  is finite and continuous at  $\varepsilon = 0$ ;  $\text{Im} \tilde{\Sigma}_\sigma(\varepsilon + i0)$  is also finite and continuous at  $\varepsilon = 0$ . Since  $\text{Re} \tilde{\Sigma}_\sigma(+i0) - \mu = 0$  and  $\text{Re} \tilde{I}(+i0) = 0$  because of the particle-hole symmetry, it follows from either Eq. (3.1a) or (3.2) that  $\rho(0) > 0$ . Since  $\tilde{\Sigma}_\sigma(\varepsilon + i0)$  is finite and continuous at  $\varepsilon = 0$ , according to the mapping condition of Eq. (2.14d),  $\tilde{\Delta}(\varepsilon) = (-1/\pi) \text{Im} \tilde{I}(\varepsilon + i0)$  is nonzero, finite, and continuous at  $\varepsilon = 0$ ;  $\text{Re} \tilde{I}(\varepsilon + i0)$  is also finite and continuous at  $\varepsilon = 0$ . Then,  $\rho(\varepsilon) > 0$  for  $\varepsilon \simeq 0$ . Neither possibility,  $\text{Im} \tilde{\Sigma}_\sigma(+i0) < 0$  nor  $\text{Im} \tilde{\Sigma}_\sigma(+i0) = 0$ , can be excluded. According to the correspondence to the mapped Anderson model, any solution characterized by  $\text{Im} \tilde{\Sigma}_\sigma(+i0) = 0$  is one for a normal Fermi liquid at  $T = 0$  K and any solution characterized by  $\text{Im} \tilde{\Sigma}_\sigma(+i0) < 0$  is one for a normal Fermi liquid at  $T > 0$  K. If  $\text{Im} \tilde{\Sigma}_\sigma(+i0) = 0$ , it follows from Eq. (3.1a) that  $\rho(0) = \rho_0(0)$  for the Hubbard model and it follows from Eq. (3.2) that  $\rho(0) = 1/[\pi \tilde{\Delta}(0)]$  for the Anderson model, as shown by Eq. (2.27); thus,  $\tilde{\Delta}(0) = 1/[\pi \rho_0(0)]$ . Neither  $\rho(0)$  nor  $\tilde{\Delta}(0)$  is renormalized by  $U$ . As discussed in the following paragraphs, a gap inevitably opens if  $\text{Re} \tilde{\Sigma}_\sigma(\varepsilon + i0)$  is discontinuous or divergent at  $\varepsilon = 0$ . Thus, the possible solution examined in this paragraph is the only possible one for a nongapped phase under  $S^3A$ , so that it is none other than the metallic solution obtained under DMFT for a sufficiently small  $U$ , i.e.,  $U \lesssim W$ .<sup>(23),(24),(25)</sup> This characterization is never inconsistent with the feature of the metallic phase under DMFT,<sup>(23),(24),(25)</sup> or rather consistent with it; for example, the constancy of  $\rho(0) = \rho_0(0)$  and  $\tilde{\Delta}(0) = 1/[\pi \rho_0(0)]$  at  $T = 0$  K as a function of  $U$  is realized in DMFT.<sup>(23),(24),(25)</sup>

Second, we consider one in which  $\text{Re}\tilde{\Sigma}_\sigma(\varepsilon + i0)$  remains finite as  $\varepsilon \rightarrow \pm 0$  but it is discontinuous at  $\varepsilon = 0$ . It follows from Eq. (3.3) that

$$\lim_{\varepsilon \rightarrow \pm 0} \text{Im}\tilde{\Sigma}_\sigma(\varepsilon + i0) = -\infty. \quad (3.4)$$

Since  $\text{Im}\tilde{\Sigma}_\sigma(\varepsilon + i0)$  is nonzero at least for a sufficiently small  $|\varepsilon|$ , if this type of solution is possible, no complete gap but only a pseudo-gap or a zerogap opens.

Third, we consider one in which  $\text{Re}\tilde{\Sigma}_\sigma(\varepsilon + i0)$  as well as  $\text{Im}\tilde{\Sigma}_\sigma(\varepsilon + i0)$  continuously diverges as  $\varepsilon \rightarrow 0$ :

$$\lim_{\varepsilon \rightarrow -0} \text{Re}\tilde{\Sigma}_\sigma(\varepsilon + i0) = -\infty, \quad \lim_{\varepsilon \rightarrow +0} \text{Re}\tilde{\Sigma}_\sigma(\varepsilon + i0) = +\infty, \quad (3.5)$$

as well as Eq. (3.4). If this type of solution is possible, a zerogap opens in it.

Fourth, we consider one in which  $\text{Re}\tilde{\Sigma}_\sigma(\varepsilon + i0)$  discontinuously diverges at  $\varepsilon = 0$ , i.e.,  $\text{Re}\tilde{\Sigma}_\sigma(\varepsilon + i0)$  includes the delta function  $\delta(\varepsilon)$ . In this case,  $\tilde{\Sigma}_\sigma(\varepsilon + i0)$  cannot be analytical in the upper-half plane. No self-consistent solution of this type is possible.

Last, we consider a possible solution in which  $\text{Re}\tilde{\Sigma}_\sigma(\varepsilon + i0)$  continuously diverges as  $\varepsilon \rightarrow \pm 0$  and  $\text{Im}\tilde{\Sigma}_\sigma(\varepsilon + i0)$  discontinuously diverges at  $\varepsilon = 0$ , i.e.,  $\tilde{\Sigma}_\sigma(\varepsilon + i0)$  has a pole at  $\varepsilon = 0$  such that

$$\tilde{\Sigma}_\sigma(\varepsilon + i0) - \mu = \frac{\tilde{\lambda}_\Sigma}{\varepsilon + i0} + \dots. \quad (3.6)$$

Then,  $\rho(0) = 0$  in this type of solution. A necessary condition for a complete gap as large as  $\epsilon_g$  to open in the Hubbard model is that  $\tilde{\Sigma}_\sigma(\varepsilon + i0)$  has a pole at  $\varepsilon = 0$ , the residue of the pole is large enough, as discussed below, and  $\text{Im}\tilde{\Sigma}_\sigma(\varepsilon + i0) = 0$  for  $|\varepsilon| < \epsilon_g/2$  but  $\varepsilon \neq 0$ . There is a zero point  $z_E$  such that it satisfies

$$\varepsilon + \mu - z_E - \text{Re}\tilde{\Sigma}_\sigma(\varepsilon + i0) = 0, \quad (3.7)$$

for any finite  $\varepsilon$  but  $\varepsilon \neq 0$ . Since  $\rho_0(E) > 0$  for  $|E| < W/2$  and  $\rho_0(E) = 0$  for  $|E| > W/2$ ,  $|\varepsilon + \mu - \text{Re}\tilde{\Sigma}_\sigma(\varepsilon + i0)| > W/2$  for any  $\varepsilon$  of  $|\varepsilon| < \epsilon_g/2$  has to be satisfied for a complete gap as large as  $\epsilon_g$  to open. Thus, the bandwidth  $W$  has to be finite and the residue  $\tilde{\lambda}_\Sigma$  has to be as large as<sup>\*)</sup>

$$\tilde{\lambda}_\Sigma > \frac{1}{4} (\epsilon_g^2 + \epsilon_g W). \quad (3.8)$$

If the bandwidth  $2W$  is finite, a possibility that a complete gap opens can never be excluded. This type of solution is the only possible one for a complete-gapped phase under  $S^3A$ , so that it is none other than the insulating solution obtained under DMFT for a sufficiently large  $U$ , i.e.,  $U \gtrsim W$ .<sup>23), 24), 25)</sup> This characterization is not inconsistent with the feature of the gapped phase under DMFT,<sup>23), 24), 25)</sup> or rather consistent with it, as discussed below.

In this subsection, we consider an unspecified finite bandwidth model. Any possibility of nongapped, zerogapped, and complete-gapped phases cannot be excluded

---

<sup>\*)</sup> If the bare bandwidth is infinite such as  $W = +\infty$ , it is easy to see from Eqs. (3.1a) and (3.7) that no complete gap but only a zerogap can open.

within the qualitative analysis of this section. However, it is certain that  $\text{Im}\tilde{\Sigma}_\sigma(+i0)$  is divergent in any possible type of gapped phase. Thus, it can be concluded on the basis of the analysis in §2.3.2 that  $T_K = 0$  K or  $T_K = +0$  K in any gapped phase and that the residual entropy or entropy at  $T = +0$  K per unit cell is therefore nonzero in it. As regards a complete-gapped phase, in particular, the entropy at  $T = +0$  K is  $k_B \ln 2$  per unit cell, as studied in Appendix B.

In any possible solution of a zerogap or complete-gap phase, inevitably  $\tilde{\Delta}(0) = 0$  under  $S^3A$ , as studied in Appendix A, as well as  $\rho(0) = 0$ . The feature that  $\rho(0) = 0$  and  $\tilde{\Delta}(0) = 0$  in the possible gapped phase is different from the feature that either  $\rho(0) = \rho_0(0)$  or  $\tilde{\Delta}(0) = 1/[\pi\rho_0(0)]$  is nonzero and constant in the possible nongapped phase; therefore,  $\pi\tilde{\Delta}(0)\rho(0) = 1$  in it. If a metal-insulator transition occurs at  $T = 0$  K or  $T = +0$  K under  $S^3A$ , not only the residual entropy or entropy at  $T = +0$  K, but also  $\rho(0)$  and  $\tilde{\Delta}(0)$  jump at the transition. The jumps of  $\rho(0)$  and  $\tilde{\Delta}(0)$  are consistent with those in the metal-insulator transition in DMFT.<sup>(23),(24),(25)</sup>

#### §4. Gapped phase beyond $S^3A$

In this section, we consider only possible complete-gapped phases at  $T = 0$  K or  $T = +0$  K beyond  $S^3A$  in a finite or infinite bandwidth model. A precise model for  $\rho_0(\varepsilon)$  is neither specified.

The single-site Green function of the Hubbard model is given by  $R_\sigma(\varepsilon + i0)$  defined by Eq. (2.18) and the local Green function of the mapped Anderson model is given by Eq. (2.12), i.e., by

$$\tilde{G}_\sigma(\varepsilon + i0) = \frac{1}{\varepsilon + i0 - \tilde{S}(\varepsilon + i0)}, \quad (4.1)$$

where

$$\tilde{S}(\varepsilon + i0) = [\tilde{\Sigma}_\sigma(\varepsilon + i0) - \mu] + \tilde{I}(\varepsilon + i0). \quad (4.2)$$

If the multisite  $\Delta\Sigma_\sigma(\varepsilon + i0, \mathbf{k})$  is normal or anomalous,  $\tilde{\Sigma}_\sigma(\varepsilon + i0)$  and  $\tilde{I}(\varepsilon + i0)$ , i.e.,  $\tilde{S}(\varepsilon + i0)$  of the mapped Anderson model is determined in a way such that  $R_\sigma(\varepsilon + i0)$  and  $\tilde{G}_\sigma(\varepsilon + i0)$  are equal to each other:  $R_\sigma(\varepsilon + i0) = \tilde{G}_\sigma(\varepsilon + i0)$ , as shown in Eq. (2.14b). In all possible solutions, it follows that

$$R_\sigma(\varepsilon + i0) = -\frac{1}{\pi} \int_{-\infty}^{+\infty} dx \frac{\rho(x)}{\varepsilon + i0 - x}, \quad (4.3)$$

where  $\rho(\varepsilon) = (-1/\pi)\text{Im}R_\sigma(\varepsilon + i0)$  is the density of states, and

$$\tilde{S}(\varepsilon + i0) = -\frac{1}{\pi} \int_{-\infty}^{+\infty} dx \frac{\text{Im}\tilde{S}(x + i0)}{\varepsilon + i0 - x}. \quad (4.4)$$

We assume that a complete gap as large as  $\epsilon_g$  opens in the  $\rho(\varepsilon)$  of a self-consistent solution:  $\rho(\varepsilon) = 0$  for  $|\varepsilon| < \epsilon_g/2$ . Then, Eq. (4.3) can be described as

$$R_\sigma(\varepsilon + i0) = -\frac{1}{\pi} \left( \int_{-\infty}^{-\epsilon_g/2} dx + \int_{+\epsilon_g/2}^{+\infty} dx \right) \frac{\rho(x)}{\varepsilon + i0 - x}. \quad (4.5)$$

Then, for  $|\varepsilon| < \epsilon_g/2$ ,  $R_\sigma(\varepsilon+i0)$  is real and  $(d/d\varepsilon)R_\sigma(\varepsilon+i0) < 0$ . Since  $\text{Re}R_\sigma(+i0) = 0$  and  $\text{Re}\tilde{S}_\sigma(+i0) = 0$  because of the particle-hole symmetry, it is easy to see that  $1/R_\sigma(\varepsilon+i0)$  has a pole at  $\varepsilon = 0$  and that, therefore,  $\tilde{S}(\varepsilon+i0)$  also has a pole at  $\varepsilon = 0$ . If  $R_\sigma(\varepsilon+i0)$  is given by Eq. (4.5), it follows from Eq. (4.1) that  $\text{Im}\tilde{S}(\varepsilon+i0) = 0$  for  $|\varepsilon| < \epsilon_g/2$  but  $\varepsilon \neq 0$ . Thus, Eq. (4.4) can be described as

$$\tilde{S}(\varepsilon+i0) = \frac{\tilde{\lambda}_S}{\varepsilon+i0} + \delta\tilde{S}(\varepsilon+i0), \quad (4.6)$$

where  $\tilde{\lambda}_S = -1/[(d/d\varepsilon)R_\sigma(\varepsilon+i0)] > 0$  and

$$\delta\tilde{S}(\varepsilon+i0) = -\frac{1}{\pi} \left( \int_{-\infty}^{-\epsilon_g/2} dx + \int_{+\epsilon_g/2}^{+\infty} dx \right) \frac{\tilde{A}_S(x)}{\varepsilon+i0-x}, \quad (4.7)$$

where  $\tilde{A}_S(x) = \tilde{A}_S(-x) \geq 0$ . In order for a gap as large as  $\epsilon_g$  to open in the Anderson model,  $\tilde{S}(\varepsilon+i0)$  has to have a pole at  $\varepsilon = 0$  and the residue  $\tilde{\lambda}_S$  has to be so large that  $|\varepsilon - \text{Re}S_\sigma(\varepsilon+i0)|$  cannot be zero for any  $\varepsilon$  of  $|\varepsilon| < \epsilon_g/2$ ; at least,<sup>\*)</sup>

$$\tilde{\lambda}_S > \frac{1}{4} \left[ \epsilon_g^2 + 2\epsilon_g |\delta\tilde{S}(\epsilon_g/2+i0)| \right], \quad (4.8)$$

has to be satisfied.

Since  $\tilde{S}(\varepsilon+i0)$  is the sum of  $\tilde{\Sigma}_\sigma(\varepsilon+i0) - \mu$  and  $\tilde{I}(\varepsilon+i0)$ , as shown in Eq. (4.2), it follows that

$$\tilde{\Sigma}_\sigma(\varepsilon+i0) - \mu = \frac{\tilde{\lambda}_\Sigma}{\varepsilon+i0} + \delta\tilde{S}_\Sigma(\varepsilon+i0), \quad \tilde{I}(\varepsilon+i0) = \frac{\tilde{\lambda}_\Gamma}{\varepsilon+i0} + \delta\tilde{S}_\Gamma(\varepsilon+i0), \quad (4.9)$$

where  $\tilde{\lambda}_\Sigma + \tilde{\lambda}_\Gamma = \tilde{\lambda}_S$ , and that

$$\delta\tilde{S}_\Sigma(\varepsilon+i0) = -\frac{1}{\pi} \left( \int_{-\infty}^{-\epsilon_g/2} dx + \int_{+\epsilon_g/2}^{+\infty} dx \right) \frac{\tilde{A}_\Sigma(x)}{\varepsilon+i0-x}, \quad (4.10)$$

$$\delta\tilde{S}_\Gamma(\varepsilon+i0) = -\frac{1}{\pi} \left( \int_{-\infty}^{-\epsilon_g/2} dx + \int_{+\epsilon_g/2}^{+\infty} dx \right) \frac{\tilde{A}_\Gamma(x)}{\varepsilon+i0-x}, \quad (4.11)$$

where  $\tilde{A}_\Sigma(x) + \tilde{A}_\Gamma(x) = \tilde{A}_S(x)$  for  $|x| > \epsilon_g/2$ ; in general,  $\tilde{\lambda}_\Sigma \geq 0$ ,  $\tilde{\lambda}_\Gamma \geq 0$ ,  $\tilde{A}_\Sigma(x) = \tilde{A}_\Sigma(-x) \geq 0$  for  $|x| > \epsilon_g/2$ , and  $\tilde{A}_\Gamma(x) = \tilde{A}_\Gamma(-x) = \tilde{\Delta}(x) \geq 0$  for  $|x| > \epsilon_g/2$ .

An issue is whether or not  $\tilde{I}(\varepsilon+i0)$  can have a pole at  $\varepsilon = 0$ , i.e., whether or not  $\tilde{\lambda}_\Gamma$  can be nonzero. There are two possible cases for  $\tilde{\lambda}_\Gamma$ :  $\tilde{\lambda}_\Gamma = 0$  and  $\tilde{\lambda}_\Gamma > 0$ . First, we examine the case in which  $\tilde{\lambda}_\Gamma = 0$ , i.e.,  $\tilde{\lambda}_\Sigma > 0$  and the single-site  $\tilde{\Sigma}_\sigma(\varepsilon+i0)$  has a pole at  $\varepsilon = 0$ . According to the analysis in Appendix A,  $\tilde{\lambda}_\Gamma = 0$  means that the

---

<sup>\*)</sup> If the multisite  $\Delta\Sigma_\sigma(\varepsilon+i0, \mathbf{k})$  is continuous as a function of  $\mathbf{k}$ , Eq. (3.8) or a similar one has to be satisfied for a complete gap to open in the Hubbard model, so that no complete gap can open if  $W$  is infinite. If it can be discontinuous as a function of  $\mathbf{k}$ , a complete gap can open even if  $W$  is infinite, as examined for a continuous model in Appendix C. If a complete gap opens in a finite or infinite bandwidth model, Eq. (4.8) has to be satisfied in either model.

multisite  $\Delta\Sigma_\sigma(\varepsilon + i0, \mathbf{k})$  has no pole at  $\varepsilon = 0$ . The possibility of this type of solution cannot be excluded within the qualitative analysis in this section.

Second, we examine the case in which  $\tilde{\lambda}_\Gamma > 0$ , i.e.,  $\tilde{\Delta}(\varepsilon + i0)$  includes the delta function  $\delta(\varepsilon)$ . In the original definition of  $\tilde{F}(\varepsilon + i0)$ , which is given by Eq. (2.13a), we first assume that

$$\tilde{\Delta}(\varepsilon) = \tilde{\lambda}_\Gamma \frac{\tilde{\gamma}_\Gamma}{\varepsilon^2 + \tilde{\gamma}_\Gamma^2}, \quad (4.12)$$

where  $\tilde{\gamma}_\Gamma$  is nonzero and finite such that  $0 < \tilde{\gamma}_\Gamma \ll \epsilon_g$ ; then we eventually take the limit of  $\tilde{\gamma}_\Gamma \rightarrow 0$ . It is evident that  $T_K$  is neither zero nor infinitesimal if  $\tilde{\gamma}_\Gamma$  is neither zero nor infinite. If  $T_K$  defined by Eq. (2.25) is used, it follows from Eq. (2.26) that

$$\tilde{\Sigma}_\sigma(\varepsilon + i0) = \frac{1}{2}U + \left[1 - \pi\tilde{\Delta}(0)/(2\tilde{W}_s k_B T_K)\right] \varepsilon + \cdots, \quad (4.13)$$

where  $\tilde{W}_s$  is the Wilson ratio defined by Eq. (2.30). This expansion is relevant for  $|\varepsilon| \lesssim \min(\tilde{\gamma}_\Gamma, k_B T_K)$ . It follows that

$$\lim_{\tilde{\gamma}_\Gamma \rightarrow 0} \left[ \frac{d}{d\varepsilon} \text{Re} \tilde{\Sigma}_\sigma(\varepsilon + i0) \right]_{\varepsilon=0} = \lim_{\tilde{\gamma}_\Gamma \rightarrow 0} \left[ 1 - \frac{\pi\tilde{\lambda}_\Gamma}{2\tilde{W}_s k_B T_K \tilde{\gamma}_\Gamma} \right] = -\infty. \quad (4.14a)$$

The Kondo temperature  $T_K$  is a function of  $\tilde{\gamma}_\Gamma$ . There are two possible cases for  $T_K$  in the limit of  $\tilde{\gamma}_\Gamma \rightarrow 0$ :  $k_B T_K \geq \tilde{\gamma}_\Gamma$  and  $k_B T_K < \tilde{\gamma}_\Gamma$ . If  $k_B T_K \geq \tilde{\gamma}_\Gamma$ , then

$$\lim_{\tilde{\gamma}_\Gamma \rightarrow 0} \text{Re} [\tilde{\Sigma}_\sigma(-\tilde{\gamma}_\Gamma + i0) - \tilde{\Sigma}_\sigma(\tilde{\gamma}_\Gamma + i0)] = \pi\tilde{\lambda}_\Gamma/(\tilde{W}_s k_B T_K) > 0. \quad (4.14b)$$

If  $k_B T_K < \tilde{\gamma}_\Gamma$ , then

$$\lim_{\tilde{\gamma}_\Gamma \rightarrow 0} \text{Re} [\tilde{\Sigma}_\sigma(-k_B T_K + i0) - \tilde{\Sigma}_\sigma(k_B T_K + i0)] = \lim_{\tilde{\gamma}_\Gamma \rightarrow 0} \pi\tilde{\lambda}_\Gamma/(\tilde{W}_s \tilde{\gamma}_\Gamma) = +\infty. \quad (4.14c)$$

According to Eq. (4.14), the real part of  $[\tilde{\Sigma}_\sigma(\varepsilon + i0)]_{\tilde{\gamma}_\Gamma \rightarrow 0}$  is discontinuous at  $\varepsilon = 0$ , so that its imagery part is divergent at  $\varepsilon = 0$  in any possible solution of this type. If it continuously diverges as  $\varepsilon \rightarrow 0$  such that

$$\lim_{\varepsilon \rightarrow 0} [\text{Im} \tilde{\Sigma}_\sigma(\varepsilon + i0)]_{\tilde{\gamma}_\Gamma \rightarrow 0} = -\infty, \quad (4.15)$$

a zerogap opens, which contradicts the opening of a complete gap. A complete gap opens only if it discontinuously diverges at  $\varepsilon = 0$  such that

$$[\text{Im} \tilde{\Sigma}_\sigma(\varepsilon + i0)]_{\tilde{\gamma}_\Gamma \rightarrow 0} \propto -\delta(\varepsilon), \quad (4.16)$$

i.e., only if  $\tilde{\Sigma}_\sigma(\varepsilon + i0)$  has a pole at  $\varepsilon = 0$ . This is only possible if  $T_K \rightarrow 0\text{K}$  in the limit of  $\tilde{\gamma}_\Gamma \rightarrow 0$ . Since  $\tilde{\Delta}(0) \rightarrow +\infty$  as  $\tilde{\gamma}_\Gamma \rightarrow 0$ , i.e.,  $\tilde{\Delta}(0) > 0$ , it is probable that  $T_K = +0\text{K}$  rather than  $T_K = 0\text{K}$ ; at least, the entropy at  $T = +0\text{K}$  is nonzero. This type of solution is only possible beyond  $S^3\text{A}$ , as studied in Appendix A.

In this section, we list all the possible types of solution for a complete-gapped phase beyond  $S^3\text{A}$  in an unspecified model. It is shown that  $\tilde{\Sigma}_\sigma(\varepsilon + i0)$  has to have a pole at  $\varepsilon = 0$  in any of them, which means that  $T_K = 0\text{K}$  or  $T_K = +0\text{K}$  in it. Thus, the residual entropy or entropy at  $T = +0\text{K}$  is nonzero in it; the entropy at  $T = +0\text{K}$  is  $k_B \ln 2$  per unit cell, as studied in Appendix B.



### §5. Discussion

Since the density of states  $\rho(\varepsilon)$  is independent of wave number, it is essentially a single-site or local property in any model. It is therefore reasonably anticipated that  $\rho(\varepsilon)$  can be described by that of a *local or impurity* model even for any model, even beyond S<sup>3</sup>A or DMFT; S<sup>3</sup>A and DMFT are exactly equivalent to each other. According to the Kondo-lattice theory,<sup>17),18),19),26)</sup> the  $\rho(\varepsilon)$  of the Hubbard model can be actually described by that of the *impurity* Anderson model, either under or beyond S<sup>3</sup>A, as shown in Eq. (2·19b). Since the site-diagonal  $R_\sigma(i\varepsilon_l)$  includes the total  $\Sigma_\sigma(i\varepsilon_l, \mathbf{k})$ , as shown in Eq. (2·18), the mapping condition of Eq. (2·14d) depends on the multisite  $\Delta\Sigma_\sigma(i\varepsilon_l, \mathbf{k})$ , i.e., the hybridization energy  $\tilde{\Delta}(\varepsilon)$  of the Anderson model to be mapped depends on  $\Delta\Sigma_\sigma(i\varepsilon_l, \mathbf{k})$ ; the multisite  $\Delta\Sigma_\sigma(i\varepsilon_l, \mathbf{k})$  is approximately or rigorously considered beyond S<sup>3</sup>A, although it is ignored under S<sup>3</sup>A. Beyond S<sup>3</sup>A, therefore, all of the  $\tilde{\Delta}(\varepsilon)$ ,  $\tilde{\Sigma}_\sigma(i\varepsilon_l)$ , and  $\Delta\Sigma_\sigma(i\varepsilon_l, \mathbf{k})$  have to be self-consistently calculated with each other to satisfy the mapping condition of Eq. (2·14d). The effects of intersite correlations, which are described in terms of the multisite  $\Delta\Sigma_\sigma(i\varepsilon_l, \mathbf{k})$  of the Hubbard model, can be described in terms of the self-consistent  $\tilde{\Delta}(\varepsilon)$  of the Anderson model, at least, as regards any local property such as  $\rho(\varepsilon)$ .

The residual entropy or entropy at  $T = +0$  K is nonzero for any possible complete-gapped phase either under or beyond S<sup>3</sup>A or DMFT; in particular, the entropy at  $T = +0$  K is  $k_B \ln 2$  per unit cell, as studied in Appendix B. In Brinkman and Rice's theory,<sup>12)</sup> the metal-insulator transition appears continuous. According to the analysis in the present paper, it is discontinuous such that the residual entropy or entropy at  $T = +0$  K jumps at the transition.

Theories based on DMFT<sup>23),24),25)</sup> are consistent with Brinkman and Rice's theory except for the hysteresis in DMFT. Since the transition is discontinuous, it is reasonable that the hysteresis appears. Under DMFT, a metal-insulator transition occurs at  $U_{c2}$  as  $U$  increases, while it occurs at  $U_{c1}$ , which is smaller than  $U_{c2}$ , as  $U$  decreases. According to the analysis in the present paper, the hysteresis can be explained in the following way: the Kondo temperature  $T_K$  continuously vanishes at  $U_{c2}$  as  $U$  increases, while it discontinuously becomes nonzero at  $U_{c1}$  as  $U$  decreases; and the residual entropy or entropy at  $T = +0$  K discontinuously becomes nonzero at  $U_{c2}$  as  $U$  increases, while it discontinuously becomes zero at  $U_{c1}$  as  $U$  decreases. The jump of  $T_K$  at  $U_{c1}$  means that the internal energy also jumps at  $U_{c1}$ .

In general, an infinitely degenerate ground state is unstable even in the presence of an infinitesimal perturbation, except for a symmetry-broken state in which a continuous symmetry is broken and rigidity appears.<sup>42)</sup> This is why the third law of thermodynamics is valid. It is therefore doubtful that a gapped phase with no symmetry broken, in which the residual entropy or entropy at  $T = +0$  K is nonzero, is stable under or beyond S<sup>3</sup>A in the presence of an infinitesimal perturbation due to the communication of electrons between the Hubbard model and an electron reservoir, i.e., in the grand canonical ensemble, as reported in a previous paper.<sup>26)</sup>

In two dimensions and higher, symmetry can be broken when the temperature  $T$  is low enough and, at least, at the absolute zero Kelvin. An infinitely degenerate

ground state in which the third law of thermodynamics is broken is unstable against an ordered state such as an antiferromagnetic state. The third law of thermodynamics is valid in any ordered ground state because rigidity appears in it.<sup>42)</sup>

The Kondo temperature  $T_K$  or  $k_B T_K$  is the energy scale of single-site quantum spin fluctuations. Electrons behave like local moments in a high- $T$  phase at  $T \gg T_K$ , because single-site thermal spin fluctuations overcome single-site quantum ones. The high- $T$  phase at  $T \gg T_K$  is in a sense a paramagnetic type of the Mott insulator, which is not the ground state. If an antiferromagnetic gap opens, the ground state can also be an insulator.<sup>43)</sup> A local-moment type of antiferromagnetic insulator, in which the Néel temperature  $T_N$  is as high as  $T_N \gg T_K$ , is an antiferromagnetic type of the Mott insulator, because electrons behave like local moments in its paramagnetic phase at  $T > T_N$ . If electrons are not half-filled and the nesting of the Fermi surface is not sharp,  $T_N$  is low or it vanishes. If  $T_N$  is nonzero and  $T_N \ll T_K$ , electrons are itinerant in a paramagnetic phase at  $T_N < T \ll T_K$ , so that the magnetism at  $T \leq T_N$  is itinerant-electron magnetism. If  $T_K$  is high enough, it is possible that the ground state is superconducting.<sup>5)</sup>

The Hubbard model in one dimension is particular, because no symmetry can be broken in it,<sup>44)</sup> and because the Bethe-ansatz solution was given by Lieb and Wu.<sup>45)</sup> We denote the ground-state energy in the canonical ensemble by  $E_G(N)$  as a function of the number  $N$  of electrons. Two *chemical potentials* are defined for the ground state of  $N$  electrons in the canonical ensemble:  $\mu_+(N) = E_G(N+1) - E_G(N)$  for the addition of an electron and  $\mu_-(N) = E_G(N) - E_G(N-1)$  for the removal of an electron. A *gap* is defined by  $\epsilon_g(N) = \mu_+(N) - \mu_-(N)$ . According to the Bethe-ansatz solution,  $\epsilon_g(N) = 0$  for  $N \neq L$ , while  $\epsilon_g(N) > 0$  for  $N = L$  for any nonzero  $U$ ; i.e., a *gap* opens in the spectrum of adding or removing the whole of a single electron in the exactly half-filled case. On the other hand, the third law of thermodynamics is not broken in the ground state of the Bethe-ansatz solution. If the opening of a gap in the spectrum of adding or removing an electron in the canonical ensemble is equivalent to that in the single-particle excitation spectrum in the grand canonical ensemble, the conclusion of the present paper and the argument based on the Bethe-ansatz solution are contradictory to each other.

A possible explanation for the contradiction is that the Kondo-lattice theory, which is none other than the  $1/D$  expansion theory from infinite dimensions, is inapplicable to the opposite limit of one dimension. Since the Kondo-lattice theory is rigorous in infinite dimensions as  $S^3A$  is, it is applicable at least to  $D \rightarrow +\infty$  dimensions.<sup>\*)</sup> If the explanation is right to the point, therefore, there is a critical  $D_c$  such that the Kondo-lattice theory is applicable to  $D \geq D_c$  dimensions while it is inapplicable to  $1 \leq D < D_c$  dimensions. It is desirable to determine how large or small the critical  $D_c$  is. It is also desirable to clarify why the Kondo-lattice theory is inapplicable to  $1 \leq D < D_c$  dimensions and, in particular, why the mapping of the single-site self-energy of the Hubbard model to the local self-energy of the Anderson

---

\*) If the multisite  $\Delta\Sigma_\sigma(\varepsilon + i0, \mathbf{k})$  as well as the single-site  $\tilde{\Sigma}_\sigma(\varepsilon + i0)$  is rigorously calculated, the Kondo-lattice theory that includes no conventional Weiss mean field, if it is applicable and relevant, is rigorous for any  $D$  dimensions, but within the constraint Hilbert subspace in which no symmetry is allowed to be broken the same as  $S^3A$  for infinite dimensions.

model is impossible for  $1 \leq D < D_c$ . Even if the Kondo-lattice theory is really inapplicable to  $1 \leq D < D_c$  dimensions, a necessary condition for the opening of a gap in the single-particle excitation spectrum is that the total self-energy  $\Sigma_\sigma(\varepsilon + i0, \mathbf{k})$  has a pole at  $\varepsilon = 0$  at least for  $\mathbf{k} = \mathbf{k}_F$  on the hypersurface defined by Eq. (2.4), as discussed in the last paragraph in §2.2. Since the self-energy  $\Sigma_\sigma(\varepsilon + i0, \mathbf{k})$  having a pole at  $\varepsilon = 0$  implies or at least suggests that there are degenerate zero-energy bosonic excitations, which are responsible for the divergent imaginary part of the self-energy, it is desirable to elucidate why and how the residual entropy can be zero even in such an anomalous situation.

Another possible explanation for the contradiction is that the opening of a gap in the canonical ensemble is not equivalent to that in the single-particle excitation spectrum in the grand canonical ensemble. Although the above well-known argument by Lieb and Wu is certainly based on the exact result of  $[\epsilon_g(N)]_{N=L} > 0$ , which is obtained by the Bethe-ansatz solution, the inequality of  $\epsilon_g(N) > 0$  is not a sufficient condition for the ground state being an insulator; for example, the very same inequality is satisfied in a metallic fine particle if the long-range Coulomb interaction is explicitly considered. Thus, the possibility that the half-filled ground state is a metal is not excluded by the Bethe-ansatz solution itself. It should be examined whether a gap opens in the single-particle excitation spectrum in the grand canonical ensemble. If the Kondo-lattice theory is applicable to one dimension, a combination of the ground-state degeneracy of the Bethe-ansatz solution and the conclusion of the present paper means that no complete gap opens in the single-particle excitation spectrum. It is desirable to study the conductivity itself in the presence of electrodes, which plays a role of an electron reservoir for the Hubbard model.

In the continuous limit of  $a \rightarrow 0$ , where  $a$  is the lattice constant, the Hubbard model in one dimension is substituted for the Tomonaga-Luttinger model with umklapp terms;<sup>27), 28), 29)</sup> and the Tomonaga-Luttinger model is further substituted for a boson model, in which a charge gap opens.<sup>46)</sup> Since the Tomonaga-Luttinger model is a continuous or nonlattice model, the Kondo-lattice theory is inapplicable to it, i.e., the Kondo effect is irrelevant in it, as discussed in Appendix C. If the Kondo-lattice theory is applicable to the Hubbard model in one dimension, the opening of a complete gap in the Tomonaga-Luttinger or Boson model, in which the Kondo effect is irrelevant, is not a proof that contradicts the conclusion of the present paper for the Hubbard model, in which the Kondo effect is relevant.

## §6. Conclusion

The self-energy of the Hubbard model is decomposed into the single-site and multisite ones. According to the Kondo-lattice theory, the single-site self-energy is equal to that of the mapped Anderson model that is self-consistently determined in a way such that the density of states of the Hubbard model per unit cell is equal to that of the Anderson model. If the multisite self-energy is ignored under the supreme single-site approximation (S<sup>3</sup>A) or if it is approximately or rigorously

considered beyond  $S^3A$ , the density of states per unit cell is simply given by

$$\rho(\varepsilon) = -\frac{1}{\pi} \text{Im} \frac{1}{\varepsilon + i0 - \tilde{S}(\varepsilon + i0)}, \quad (6.1)$$

where  $\tilde{S}(\varepsilon + i0) = \tilde{\epsilon}_d - \tilde{\mu} + \tilde{\Sigma}_\sigma(\varepsilon + i0) + \tilde{F}(\varepsilon + i0)$ , where  $\tilde{\epsilon}_d$ ,  $\tilde{\mu}$ ,  $\tilde{\Sigma}_\sigma(\varepsilon + i0)$ , and  $\tilde{F}(\varepsilon + i0)$  are the level of localized electrons, the chemical potential, the self-energy, and the hybridization term, respectively, of the mapped Anderson model; and  $\tilde{\mu} - \tilde{\epsilon}_d = \mu - \epsilon_d$ , where  $\mu$  and  $\epsilon_d$  are the chemical potential and band center, respectively, of the Hubbard model. The Kondo temperature  $T_K$  or  $k_B T_K$  of the mapped Anderson model is also the low-energy scale of single-site quantum spin fluctuations in the Hubbard model.

In the exactly half-filled case of the Hubbard model on a bipartite lattice,  $\text{Re}\tilde{S}(+i0) = 0$  because of the particle-hole symmetry. A complete gap as large as  $\epsilon_g$  can open in  $\rho(\varepsilon)$ , if and only if  $\tilde{S}(\varepsilon + i0)$  has a pole at  $\varepsilon = 0$ , the residue of the pole is so large that  $|\varepsilon - \text{Re}\tilde{S}_\sigma(\varepsilon + i0)|$  cannot be zero for any  $|\varepsilon| < \epsilon_g/2$ , and  $\text{Im}\tilde{S}(\varepsilon + i0) = 0$  for any  $|\varepsilon| < \epsilon_g/2$  but  $\varepsilon = 0$ . Such an anomalous  $\tilde{S}(\varepsilon + i0)$  is possible if and only if  $T_K = 0$  K or  $T_K = +0$  K. If  $T_K = 0$  K, the residual entropy is nonzero. If  $T_K = +0$  K, the ground state is a singlet, while the entropy at  $T = +0$  K is nonzero; the entropy discontinuously vanishes at  $T = 0$  K with decreasing  $T$ . Thus, the Mott insulator with no symmetry broken, if it is possible, is characterized by nonzero residual entropy or nonzero entropy at  $T = +0$  K.

The conclusion of the present paper is consistent with Brinkman and Rice's theory, in which the specific-heat coefficient diverges as the onsite  $U$  approaches a transition point from a metallic phase and no symmetry is broken in an insulating phase; a combination of these two properties means that the residual entropy or entropy at  $T = +0$  K is nonzero in the insulating phase. It is also consistent with the appearance of hysteresis in the dynamical mean-field theory.

According to the well-known argument based on the Bethe-ansatz solution, on the other hand, the half-filled ground state in one dimension is the Mott insulator for any nonzero onsite  $U$ , although the third law of thermodynamics is not broken in it; thus, the well-known argument and the conclusion of the present paper contradict each other. A possible explanation for the contradiction is that the Kondo-lattice theory, which is none other than the  $1/D$  expansion theory from infinite dimensions, is inapplicable to one dimension. The other possible explanation is that a gap in the spectrum of adding or removing an electron in the canonical ensemble, which is given by the Bethe-ansatz solution, is different from a gap in the single-particle excitation spectrum in the grand canonical ensemble, which is studied in the present paper. Since the opening of the gap in the canonical ensemble is not a sufficient condition for the ground state being an insulator, the well-known argument is not a proof that the half-filled ground state in one dimension is the Mott insulator. If the Kondo-lattice theory is applicable to one dimension, a combination of the conclusion of the present paper and the ground-state degeneracy of the Bethe-ansatz solution means that the half-filled ground state in one dimension is not the Mott insulator. It is desirable to elucidate which explanation is true in the near future.

## Appendix A

### — Anomalies of Gapped Phases —

In order for a complete gap as large as  $\epsilon_g$  to open, the single-site  $\tilde{\Sigma}_\sigma(\varepsilon + i0)$  has to have a pole at  $\varepsilon = 0$ , as examined in §§3.2 and 4, so that the total  $\Sigma_\sigma(\varepsilon + i0, \mathbf{k})$  also has to have a pole at  $\varepsilon = 0$ . However, it does not have to have two or more poles in the gap region of  $|\varepsilon| < \epsilon_g/2$ , as discussed below.

If the total  $\Sigma_\sigma(\varepsilon + i0, \mathbf{k})$  has two or more poles in the region of  $|\varepsilon| < \epsilon_g/2$ , there is at least a pair of adjacent poles: e.g., one at  $\varepsilon = p_1$  and the other at  $\varepsilon = p_2$ , where  $-\epsilon_g/2 < p_1 < p_2 < \epsilon_g/2$ . Then, it follows that

$$\lim_{\varepsilon \rightarrow p_1+0} \text{Re} \Sigma_\sigma(\varepsilon + i0, \mathbf{k}) = +\infty, \quad \lim_{\varepsilon \rightarrow p_2-0} \text{Re} \Sigma_\sigma(\varepsilon + i0, \mathbf{k}) = -\infty. \quad (\text{A}\cdot 1)$$

Since  $\Sigma_\sigma(\varepsilon + i0, \mathbf{k})$  has no singularity between poles and is continuous between poles, there is a zero point  $z_\varepsilon$  between poles such that it satisfies

$$\text{Re}[z_\varepsilon + \mu - E(\mathbf{k}) - \Sigma_\sigma(z_\varepsilon + i0, \mathbf{k})] = 0. \quad (\text{A}\cdot 2)$$

Since the zero point  $z_\varepsilon$  is in the gap region of  $|\varepsilon| < \epsilon_g/2$ , it follows that

$$\text{Im}[z_\varepsilon + \mu - E(\mathbf{k}) - \Sigma_\sigma(z_\varepsilon + i0, \mathbf{k})] = 0. \quad (\text{A}\cdot 3)$$

Since the density of states  $\rho(\varepsilon)$  of the Hubbard model is given by Eq.(2·19) with Eq.(2·18), an in-gap state inevitably appears at each zero point  $\varepsilon = z_\varepsilon$  between each pair of adjacent poles. This conclusion contradicts the opening of a complete gap. Thus, the total self-energy  $\Sigma_\sigma(\varepsilon + i0, \mathbf{k})$  does not have to have two or more poles in the gap region of  $|\varepsilon| < \epsilon_g/2$  for a complete gap as large as  $\epsilon_g$  to open.

A complete gap as large as  $\epsilon_g$  opens in the Hubbard model if and only if single-site and multisite self-energies are given by

$$\tilde{\Sigma}_\sigma(i\varepsilon_l) - \mu = \frac{\tilde{\lambda}_\Sigma}{i\varepsilon_l} - \frac{1}{\pi} \left( \int_{-\infty}^{-\epsilon_g/2} dx + \int_{+\epsilon_g/2}^{+\infty} dx \right) \frac{\tilde{\Lambda}_\Sigma(x)}{i\varepsilon_l - x}, \quad (\text{A}\cdot 4\text{a})$$

$$\Delta\Sigma_\sigma(i\varepsilon_l, \mathbf{k}) = \frac{\lambda_{\Delta\Sigma}(\mathbf{k})}{i\varepsilon_l} - \frac{1}{\pi} \left( \int_{-\infty}^{-\epsilon_g/2} dx + \int_{+\epsilon_g/2}^{+\infty} dx \right) \frac{\Lambda_{\Delta\Sigma}(x; \mathbf{k})}{i\varepsilon_l - x}, \quad (\text{A}\cdot 4\text{b})$$

where  $\tilde{\lambda}_\Sigma + \tilde{\lambda}_{\Delta\Sigma}(\mathbf{k})$  is nonzero and large enough;  $\tilde{\lambda}_\Sigma \geq 0$  and  $\tilde{\lambda}_{\Delta\Sigma}(\mathbf{k}) \geq 0$ , in general. According to Eqs. (2·18) and (A·4),

$$R_\sigma(i\varepsilon_l) = i\varepsilon_l \left[ \frac{1}{L} \sum_{\mathbf{k}} \frac{1}{\tilde{\lambda}_\Sigma + \lambda_{\Delta\Sigma}(\mathbf{k})} \right] + O[(i\varepsilon_l)^3], \quad (\text{A}\cdot 5)$$

in the limit of  $|\varepsilon_l| \rightarrow 0$ . According to the mapping condition of Eq. (2·14d),

$$\tilde{\Delta}(\varepsilon) = (\tilde{\lambda}_\Gamma/\pi)\delta(\varepsilon), \quad (\text{A}\cdot 6)$$

for  $|\varepsilon| < \epsilon_g/2$ , where

$$\tilde{\lambda}_\Gamma = \begin{cases} \left[ \frac{1}{L} \sum_{\mathbf{k}} \left( \frac{1}{\tilde{\lambda}_\Sigma + \lambda_{\Delta\Sigma}(\mathbf{k})} - \frac{1}{\tilde{\lambda}_\Sigma} \right) \right]^{-1}, & \tilde{\lambda}_\Sigma > 0, \lambda_{\Delta\Sigma}(\mathbf{k}) \geq 0, \\ \left[ \frac{1}{L} \sum_{\mathbf{k}} \frac{1}{\lambda_{\Delta\Sigma}(\mathbf{k})} \right]^{-1}, & \tilde{\lambda}_\Sigma = 0, \lambda_{\Delta\Sigma}(\mathbf{k}) > 0. \end{cases} \quad (\text{A}\cdot 7)$$

If  $\lambda_{\Delta\Sigma}(\mathbf{k}) = 0$  or  $\Delta\Sigma_\sigma(\varepsilon + i0, \mathbf{k})$  has no pole at  $\varepsilon = 0$ , then  $\tilde{\lambda}_\Gamma = 0$ , i.e.,  $\tilde{I}(\varepsilon + i0)$  has no pole at  $\varepsilon = 0$ . If  $\lambda_{\Delta\Sigma}(\mathbf{k}) > 0$  or  $\Delta\Sigma_\sigma(\varepsilon + i0, \mathbf{k})$  has a pole at  $\varepsilon = 0$ , then  $\tilde{\lambda}_\Gamma > 0$ , i.e.,  $\tilde{I}(\varepsilon + i0)$  has a pole at  $\varepsilon = 0$ . According to the analysis in §4, if  $\tilde{I}(\varepsilon + i0)$  has a pole at  $\varepsilon = 0$ , the single-site  $\tilde{\Sigma}_\sigma(\varepsilon + i0)$  also has a pole at  $\varepsilon = 0$ . Thus, any self-consistent solution characterized by  $\tilde{\lambda}_\Sigma = 0$  and  $\lambda_{\Delta\Sigma}(\mathbf{k}) > 0$  is impossible.

Next, we assume that the single-site  $\tilde{\Sigma}_\sigma(\varepsilon + i0)$  is divergent at  $\varepsilon = 0$  such that

$$\lim_{\varepsilon \rightarrow 0} |\tilde{\Sigma}_\sigma(\varepsilon + i0)| = +\infty, \quad |\tilde{\Sigma}_\sigma(+i0)| = +\infty, \quad (\text{A}\cdot 8a)$$

while the multisite  $\Delta\Sigma_\sigma(\varepsilon + i0, \mathbf{k})$  is finite and continuous at  $\varepsilon = 0$  and

$$\text{Im}[\Delta\Sigma_\sigma(\varepsilon + i0, \mathbf{k})] = 0, \quad (\text{A}\cdot 8b)$$

for any  $|\varepsilon| < \epsilon_g/2$  and any  $\mathbf{k}$ . We define

$$\tilde{g}(\varepsilon) = \frac{1}{\varepsilon + i0 + \mu - \tilde{\Sigma}_\sigma(\varepsilon + i0)}, \quad Q_n(\varepsilon) = \frac{1}{L} \sum_{\mathbf{k}} [E(\mathbf{k}) + \Delta\Sigma_\sigma(\varepsilon + i0, \mathbf{k})]^n. \quad (\text{A}\cdot 9)$$

It is easy to see that

$$\lim_{\varepsilon \rightarrow 0} |\tilde{g}(\varepsilon)| = 0, \quad |\tilde{g}(0)| = 0, \quad (\text{A}\cdot 10)$$

$Q_n(\varepsilon)$  is real for  $|\varepsilon| < \epsilon_g/2$ , and  $Q_n(\varepsilon) = (-1)^n Q_n(-\varepsilon)$  for  $|\varepsilon| < \epsilon_g/2$ . According to Eq. (2·18),

$$\frac{1}{R_\sigma(\varepsilon + i0)} = \frac{1}{\tilde{g}(\varepsilon)} - Q_1(\varepsilon) - \tilde{g}(\varepsilon) [Q_2(\varepsilon) - Q_1^2(\varepsilon)] + O[\tilde{g}^2(\varepsilon)], \quad (\text{A}\cdot 11)$$

for a sufficiently small  $|\varepsilon|$ . According to the mapping condition of Eq. (2·14d),

$$\tilde{\Delta}(\varepsilon) = -\text{Im}\{\tilde{g}(\varepsilon)[Q_2(\varepsilon) - Q_1^2(\varepsilon)] + O[\tilde{g}^2(\varepsilon)]\}, \quad (\text{A}\cdot 12)$$

for a sufficiently small  $|\varepsilon|$ . It follows from Eqs. (A·10) and (A·12) that

$$\lim_{\varepsilon \rightarrow 0} \tilde{\Delta}(\varepsilon) = 0, \quad \tilde{\Delta}(0) = 0. \quad (\text{A}\cdot 13)$$

In particular, the multisite  $\Delta\Sigma_\sigma(\varepsilon + i0, \mathbf{k})$  vanishes under  $S^3A$ . Therefore, Eq. (A·13) has to be satisfied in any possible complete-gapped or zerogapped phase under  $S^3A$ .

**Appendix B****—— Entropy at  $T = +0$  K of Possible Complete-Gapped Phases ——**

If no symmetry is broken, the thermodynamic potential is given by<sup>47),48)</sup>

$$\Omega = -k_B T \sum_l \sum_{\mathbf{k}\sigma} e^{i\varepsilon_l 0^+} \left\{ \ln[-1/G_\sigma(i\varepsilon_l, \mathbf{k})] + G_\sigma(i\varepsilon_l, \mathbf{k}) \Sigma_\sigma(i\varepsilon_l, \mathbf{k}) \right\} + \Omega', \quad (\text{B}\cdot 1)$$

where  $\Omega'$  is the contribution of all the closed connected Feynman diagrams but with each electron line replaced by its renormalized one and is given by

$$\Omega' = k_B T \sum_\nu \sum_l \sum_{\mathbf{k}\sigma} \frac{1}{2\nu} G_\sigma(i\varepsilon_l, \mathbf{k}) \Sigma_{\nu\sigma}(i\varepsilon_l, \mathbf{k}), \quad (\text{B}\cdot 2)$$

where  $\Sigma_{\nu\sigma}(i\varepsilon_l, \mathbf{k})$  is the  $\nu$ th order self-energy in  $U$ , where only the  $U$  occurring in explicit interactions of the skeleton diagram is used to determine the order. Since the thermodynamic potential  $\Omega$  is stationary with respect to the self-energy such that  $\partial\Omega/\partial\Sigma_\sigma(i\varepsilon_l, \mathbf{k}) = 0$  in the functional derivative, the temperature dependence of  $\Sigma_\sigma(i\varepsilon_l, \mathbf{k})$  does not have to be considered in calculating entropy  $S = -\partial\Omega/\partial T$  except for that coming from  $i\varepsilon_l$  or the  $l$  sum.

As discussed by Luttinger and Ward,<sup>47),48)</sup> the first correction to the  $T = 0$  value of  $\Omega$  comes only from the difference between the  $l$  sum and what we would obtain if we replaced them by integrals in each electron line. Then, the first correction to the  $T = 0$  value of  $\Omega'$  is equal to that of

$$\begin{aligned} \bar{\Omega}' &= k_B T \sum_\nu \sum_l \sum_{\mathbf{k}\sigma} G_\sigma(i\varepsilon_l, \mathbf{k}) \Sigma_{\nu\sigma}(i\varepsilon_l, \mathbf{k}) \\ &= k_B T \sum_l \sum_{\mathbf{k}\sigma} G_\sigma(i\varepsilon_l, \mathbf{k}) \Sigma_\sigma(i\varepsilon_l, \mathbf{k}), \end{aligned} \quad (\text{B}\cdot 3)$$

so that the first correction to the  $T = 0$  value of the total  $\Omega$  is equal to that of

$$\begin{aligned} \bar{\Omega}(T) &= -k_B T \sum_l \sum_{\mathbf{k}\sigma} e^{i\varepsilon_l 0^+} \ln[-1/G_\sigma(i\varepsilon_l, \mathbf{k})] \\ &= \frac{1}{\pi} \sum_{\mathbf{k}\sigma} \int_{-\infty}^{+\infty} d\epsilon f(\epsilon) \text{Im} \ln[-\epsilon - i0 + E(\mathbf{k}) + \Sigma_\sigma(\epsilon + i0, \mathbf{k}) - \mu], \end{aligned} \quad (\text{B}\cdot 4)$$

where  $f(\epsilon) = 1/[e^{\epsilon/(k_B T)} + 1]$ ; in Eqs. (B·3) and (B·4),  $T = 0$  K is assumed in the self-energy except for the  $l$  sum.

For any possible complete-gap phase either under or beyond  $S^3A$ , it follows according to Eq. (A·4) that

$$\Sigma_\sigma(\epsilon + i0, \mathbf{k}) - \mu = \frac{\lambda_\Sigma(\mathbf{k})}{\epsilon + i0} - \frac{1}{\pi} \left( \int_{-\infty}^{-\epsilon_g/2} dx + \int_{+\epsilon_g/2}^{+\infty} dx \right) \frac{\Lambda_\Sigma(x; \mathbf{k})}{\epsilon + i0 - x}, \quad (\text{B}\cdot 5)$$

where  $\lambda_\Sigma(\mathbf{k}) = \tilde{\lambda}_\Sigma + \Delta\lambda_{\Delta\Sigma}(\mathbf{k}) > 0$  and  $\Lambda_\Sigma(x; \mathbf{k}) = \tilde{\Lambda}_\Sigma + \Delta\Lambda_{\Delta\Sigma}(x; \mathbf{k}) > 0$ . Then, it

follows that

$$\text{Im} \ln[-\varepsilon - i0 + E(\mathbf{k}) + \Sigma_\sigma(\varepsilon + i0, \mathbf{k}) - \mu] = \begin{cases} 0, & -\epsilon_g/2 < \varepsilon < 0, \\ -\pi/2, & \varepsilon = 0, \\ -\pi, & 0 < \varepsilon < +\epsilon_g/2, \end{cases} \quad (\text{B}\cdot 6)$$

for  $|\varepsilon| < \epsilon_g/2$ . The entropy at  $T = +0$  K per unit cell is as large as

$$S/L = - \lim_{T \rightarrow 0\text{K}} \frac{1}{L} \frac{\partial \bar{\Omega}(T)}{\partial T} = \lim_{T \rightarrow 0\text{K}} \frac{\partial}{\partial T} \int_0^{+\epsilon_g/2} d\epsilon f(\epsilon) = k_B \ln 2, \quad (\text{B}\cdot 7)$$

for any possible complete-gap phase either under or beyond S<sup>3</sup>A.

### Appendix C

#### —— Continuous Model in One Dimension ——

We consider a finite or infinite bandwidth model in one dimension in the continuous limit of  $a \rightarrow 0$ , with any of  $k_{\text{F}}a = \pi/2$ ,  $\hbar v_{\text{F}} = [(\partial/\partial k)E(k)]_{k=k_{\text{F}}}$ ,  $g = Ua$ , and  $La$  being kept constant;  $v_{\text{F}}$  is the Fermi velocity and  $g$  is the coupling constant. We assume the thermodynamic limit of  $La \rightarrow +\infty$ .

If  $a$  is nonzero and finite, the single-site Green function is given by

$$R_\sigma(\varepsilon + i0) = \frac{a}{2\pi} \int_{-\pi/a}^{+\pi/a} dk \frac{1}{\varepsilon + i0 - E(k) - S(\varepsilon + i0, k)}, \quad (\text{C}\cdot 1)$$

where

$$S(\varepsilon + i0, k) = \Sigma_\sigma(\varepsilon + i0, k) - \mu. \quad (\text{C}\cdot 2)$$

Since it vanishes in the continuous limit of  $a \rightarrow 0$ , the single-site self-energy is never well-defined in any continuous model. Thus, the Kondo-lattice theory is inapplicable to any continuous model, so that the Kondo effect is irrelevant in it.

When the infinite bandwidth model defined in §2.1 is considered, for example, the bare density of states per unit length is given by<sup>\*)</sup>

$$\bar{\rho}_0(E) = \frac{1}{a} \rho_0(E) = \frac{1}{\pi a} \frac{|t_\infty|}{E^2 + |t_\infty|^2} = \frac{1}{\pi \hbar v_{\text{F}}} \frac{1}{1 + [Ea/(\hbar v_{\text{F}})]^2}, \quad (\text{C}\cdot 3)$$

where  $a = +0$ . Since  $E(k)$  is a continuous function of  $k$ ,  $k$  is also a continuous function of  $E$ . Thus,  $S[\varepsilon + i0, k(E)]$  is a function of  $E$ ; it is simply denoted by  $S(\varepsilon + i0, E)$ . The density of state per unit length is given by

$$\bar{\rho}(\varepsilon) = \frac{1}{\pi} \int_{-\infty}^{+\infty} dE \bar{\rho}_0(E) \frac{1}{\varepsilon + i0 - E - S(\varepsilon + i0, E)}. \quad (\text{C}\cdot 4)$$

Because of the particle-hole symmetry,  $\text{Re}S(+i0, 0) = 0$ . It is easy to see that unless  $\text{Im}S(\varepsilon + i0, 0)$  is divergent at  $\varepsilon = 0$ , then  $\bar{\rho}(0) > 0$  and not  $\bar{\rho}(0) = 0$ .

---

<sup>\*)</sup> If  $\bar{\rho}_0(E) = e^{-\alpha|E|/(\hbar v_{\text{F}})}/(\pi \hbar v_{\text{F}})$  with  $\alpha = +0$  is assumed instead of Eq. (C·3), it is none other than the band cutoff introduced into the boson model.



We examine a possible complete-gapped phase whose gap is as large as  $\epsilon_g$ :  $\bar{\rho}(\varepsilon) = 0$  for  $|\varepsilon| < \epsilon_g/2$ .  $S(\varepsilon + i0, E)$  as a function of  $E$  may be continuous or discontinuous. First, we assume that  $S(\varepsilon + i0, E)$  is continuous as a function of  $E$ . According to an analysis similar to that in Appendix A,  $S(\varepsilon + i0, E)$  does not have to have two or more poles in the region of  $|\varepsilon| < \epsilon_g/2$  for a complete gap to open. Thus, it has to have zero or only one pole in the region, so that

$$S(\varepsilon + i0, E) = \frac{\lambda(E)}{\varepsilon + i0 - p(E)} - \frac{1}{\pi} \left( \int_{-\infty}^{-\epsilon_g/2} dx + \int_{+\epsilon_g/2}^{+\infty} dx \right) \frac{\Lambda(x; E)}{\varepsilon + i0 - x}, \quad (\text{C}\cdot 5)$$

where  $\lambda(E) = \lambda(-E) \geq 0$ ,  $p(E) = -p(-E)$ , and  $\Lambda(x; E) = \Lambda(-x; -E) > 0$ ; in particular,  $\lambda(0) > 0$  and  $p(0) = 0$ . If the pole  $p(E)$  depends on  $E$ ,  $S(\varepsilon + i0, E)$  as a function of  $E$  is discontinuous at  $E$  such that it satisfies  $p(E) = \varepsilon$ . Thus,  $p(E) = 0$  because of the constraint imposed by the assumption above. It is evident that

$$\lim_{E \rightarrow +\infty} [E + S(\varepsilon + i0, E)] = +\infty, \quad \lim_{E \rightarrow -\infty} [E + S(\varepsilon + i0, E)] = -\infty. \quad (\text{C}\cdot 6)$$

Then, there is a zero point  $z_E$  such that it satisfies

$$\text{Re}[\varepsilon - z_E - S(\varepsilon + i0, z_E)] = 0, \quad (\text{C}\cdot 7)$$

for any finite  $\varepsilon$  but  $\varepsilon = 0$ . The zero point  $z_E$  is a function of  $\varepsilon$ ; the function is denoted by  $z_E(\varepsilon)$ . If  $|\varepsilon| < \epsilon_g/2$  but  $\varepsilon \neq 0$ ,

$$\text{Im}\{\varepsilon - z_E(\varepsilon) - S[\varepsilon + i0, z_E(\varepsilon)]\} = 0. \quad (\text{C}\cdot 8)$$

Then,  $\bar{\rho}(\varepsilon) > 0$  for any  $\varepsilon$  but  $\varepsilon = 0$ . This conclusion contradicts the opening of a complete gap. If the bandwidth is infinite, no complete gap can open in  $\bar{\rho}(\varepsilon)$ ; however, a possibility that a zerogap opens in  $\bar{\rho}(\varepsilon)$  cannot be excluded.

If  $S(\varepsilon + i0, E)$  is allowed to be discontinuous as a function of  $E$ , numerous types of anomaly are possible because  $S(\varepsilon + i0, E)$  as a function of  $E$  does not need to be analytical in the upper or lower half complex plane of  $E$ ; for example,

$$S(\varepsilon + i0, E) = \frac{\lambda(E)}{\varepsilon + i0} + \delta S(\varepsilon + i0, E), \quad (\text{C}\cdot 9a)$$

where  $\lambda(E) \geq 0$  and, in particular,  $\lambda(0) > 0$  and

$$\delta S(\varepsilon + i0, E) = \begin{cases} \frac{E}{|E|} \left[ \sqrt{E^2 + (\epsilon_g/2)^2} - |E| \right], & E \neq 0, \\ 0, & E = 0, \end{cases} \quad (\text{C}\cdot 9b)$$

$$S(\varepsilon + i0, E) = \frac{(\epsilon_g/2)^2}{\varepsilon + i0 + E}, \quad (\text{C}\cdot 10)$$

and so on. If  $S(\varepsilon + i0, E)$  is given by Eq. (C·9) or (C·10), a complete gap as large as  $\epsilon_g$  opens, even if the bandwidth is infinite. A possibility that a complete gap opens

in  $\bar{\rho}(\varepsilon)$  cannot be excluded. A necessary condition for a complete gap to open is that  $S(\varepsilon + i0, E = 0)$  or  $S(\varepsilon + i0, k = \pm k_F)$  has a pole at  $\varepsilon = 0$  and  $S[\varepsilon + i0, E(k)]$  is discontinuous as a function of  $E$  or  $k$ .

The single-band model examined above can be substituted for a two-band model. We define two types of creation and annihilation operators by

$$\alpha_{k\sigma}^\dagger = d_{(k+k_F)\sigma}^\dagger, \quad \alpha_{k\sigma} = d_{(k+k_F)\sigma}, \quad \beta_{k\sigma}^\dagger = d_{(k-k_F)\sigma}^\dagger, \quad \beta_{k\sigma} = d_{(k-k_F)\sigma}, \quad (\text{C}\cdot 11)$$

for  $-k_F < k < k_F$ ;  $\alpha$  and  $\beta$  electrons correspond to right-going and left-going ones, respectively, in the single-band model. The one-body term is substituted for

$$\mathcal{H}_0 = \sum_{k\sigma} \hbar v_F k \left( \alpha_{k\sigma}^\dagger \alpha_{k\sigma} - \beta_{k\sigma}^\dagger \beta_{k\sigma} \right). \quad (\text{C}\cdot 12)$$

The two-body term is decomposed into several terms according to the numbers of  $\alpha_{k\sigma}^\dagger$ ,  $\alpha_{k\sigma}$ ,  $\beta_{k\sigma}^\dagger$ , and  $\beta_{k\sigma}$ , some of which are the so called umklapp terms. This two-band model is none other than the Tomonaga-Luttinger model with umklapp terms.<sup>27), 28), 29)</sup> Since the two-band model is a continuous model the same as the original single-band model, the Kondo-lattice theory is inapplicable to it. Since each of the numbers of  $\alpha$  and  $\beta$  electrons is not a conserved quantity because of the two-body term, the Green function is given, in general, by

$$\mathcal{G}_\sigma(i\varepsilon_l, k) = \begin{pmatrix} i\varepsilon_l - \hbar v_F k - S_{\alpha\alpha\sigma}(i\varepsilon_l, k) & -S_{\alpha\beta\sigma}(i\varepsilon_l, k) \\ -S_{\beta\alpha\sigma}(i\varepsilon_l, k) & i\varepsilon_l + \hbar v_F k - S_{\beta\beta\sigma}(i\varepsilon_l, k) \end{pmatrix}^{-1}. \quad (\text{C}\cdot 13)$$

If the correspondence to the original single-band model is seriously considered, the interband terms have to vanish such that

$$S_{\alpha\beta\sigma}(i\varepsilon_l, k) = S_{\beta\alpha\sigma}(i\varepsilon_l, k) = 0, \quad (\text{C}\cdot 14a)$$

and the intraband terms have to satisfy

$$S_{\alpha\alpha\sigma}(i\varepsilon_l, k) = S_{\beta\beta\sigma}(i\varepsilon_l, -k) = \Sigma_\sigma[i\varepsilon_l, E(k + k_F)] - \mu. \quad (\text{C}\cdot 14b)$$

The density of states of this two-band model is equal to that of the original single-band model, provided that Eq. (C·14) is satisfied. Thus, a necessary condition for a complete gap to open is that  $S_{\alpha\alpha\sigma}(\varepsilon + i0, k)$  has a pole at  $\varepsilon = 0$  at least for  $k = 0$  and  $S_{\alpha\alpha\sigma}(\varepsilon + i0, k)$  is discontinuous as a function of  $k$ .

It is desirable to elucidate how discontinuous  $S_{\alpha\alpha\sigma}(\varepsilon + i0, k)$  is as a function of  $k$  in the gapped phase of the Tomonaga-Luttinger model, which corresponds to the gapped phase of the boson model.<sup>46)</sup> Since it is surprising that the self-energy as a function of wave number  $k$  is discontinuous, it is desirable to elucidate what is actually responsible for the discontinuity. It is also desirable to elucidate whether the residual entropy is zero or nonzero in the gapped phases; the divergence of  $\text{Im}S_{\alpha\alpha\sigma}(+i0, 0)$  implies or at least suggests that there are degenerate zero-energy bosonic excitations, which are responsible for the divergent  $\text{Im}S_{\alpha\alpha\sigma}(+i0, 0)$ .

## References

- 1) N. F. Mott, *Metal-Insulator Transition* (London New York Philadelphia, 1990).
- 2) J. G. Bednortz and K. A. Müller, Z. Phys. B **64** (1986), 189.
- 3) P. W. Anderson, Science **235** (1987), 1196.
- 4) F. J. Ohkawa, J. Phys. Soc. Jpn. **56** (1987), 2267.
- 5) F. J. Ohkawa, J. Phys. Soc. Jpn. **78** (2009), 084712.
- 6) J. Hubbard, Proc. R. Soc. London Ser. A **276** (1963), 238.
- 7) J. Hubbard, Proc. R. Soc. London Ser. A **281** (1964), 401.
- 8) M. C. Gutzwiller, Phys. Rev. Lett. **10** (1963), 159.
- 9) M. C. Gutzwiller, Phys. Rev. **134** (1964), A923.
- 10) M. C. Gutzwiller, Phys. Rev. **137** (1965), A1726.
- 11) F. J. Ohkawa, J. Phys. Soc. Jpn. **58** (1989), 4156.
- 12) W. F. Brinkman and T. M. Rice, Phys. Rev. B **2** (1970), 4302.
- 13) W. Metzner and D. Vollhardt, Phys. Rev. Lett. **62** (1989), 324.
- 14) E. Müller-Hartmann, Z. Phys. B **74** (1989), 507.
- 15) E. Müller-Hartmann, Z. Phys. B **76** (1989), 211.
- 16) V. Janis, Z. Phys. B **83** (1991), 227.
- 17) F. J. Ohkawa, Phys. Rev. B **44** (1991), 6812.
- 18) F. J. Ohkawa, J. Phys. Soc. Jpn. **60** (1991), 3218.
- 19) F. J. Ohkawa, J. Phys. Soc. Jpn. **61** (1992), 1615.
- 20) A. Georges and G. Kotliar, Phys. Rev. B **45** (1992), 6479.
- 21) A. Georges, G. Kotliar, W. Krauth and M. J. Rosenberg, Rev. Mod. Phys. **68** (1996), 13.
- 22) Y. Kakehashi and P. Fulde, Phys. Rev. B **69** (2004), 045101.
- 23) G. Kotliar, S. Murthy and M. J. Rozenberg, Phys. Rev. Lett. **89** (2002), 046401.
- 24) G. Moeller, Q. Si, G. Kotliar, M. Rozenberg and D. S. Fisher, Phys. Rev. Lett. **74** (1995), 2082.
- 25) R. Bulla, Phys. Rev. Lett. **83** (1999), 136.
- 26) F. J. Ohkawa and T. Toyama, J. Phys. Soc. Jpn. **78** (2009), 124707.
- 27) S. Tomonaga, Prog. Theor. Phys. **5** (1950), 544.
- 28) J. M. Luttinger, J. Math. Phys. **4** (1963), 1154.
- 29) D. C. Mattis and E. H. Lieb, J. Math. Phys. **6** (1965), 304.
- 30) K. Yosida, Phys. Rev. **147** (1966), 223.
- 31) P. W. Anderson, J. of Phys. C **3** (1970), 2436.
- 32) K. G. Wilson, Rev. Mod. Phys. **47** (1975), 773.
- 33) P. Nozières, J. Low. Temp. Phys. **17** (1974), 31.
- 34) K. Yamada, Prog. Theor. Phys. **53** (1975), 970.
- 35) K. Yamada and K. Yosida, Prog. Theor. Phys. **53** (1975), 1286.
- 36) H. Shiba, Prog. Theor. Phys. **54** (1975), 967.
- 37) N. Andrei, K. Furuya and J. H. Lowenstein, Rev. Mod. Phys. **55** (1983), 331.
- 38) A. M. Tsvelik and P. B. Wiegmann, Adv. Phys. **32** (1983), 453.
- 39) A. Okiji and N. Kawakami, *Springer Series in Solid State Science* (Springer, Berlin, 1988), Vol. 77, p.63.
- 40) P. Schlottmann, Phys. Rep. **118** (1989), 1.
- 41) P. W. Anderson, Phys. Rev. **124** (1961), 41.
- 42) P. W. Anderson, *Basic Notions of Condensed Matter Physics*, Frontiers in Physics (Benjamin/Cummings, New York, 1984).
- 43) J. C. Slater, Phys. Rev. **82** (1951), 538.
- 44) N. D. Mermin and H. Wagner, Phys. Rev. Lett. **17** (1966), 1133.
- 45) E. H. Lieb and F. Y. Wu, Phys. Rev. Lett. **20** (1968), 1445.
- 46) J. Sólyom, Adv. Phys. **28** (1979), 201.
- 47) J. M. Luttinger, Phys. Rev. **119** (1960), 1153.
- 48) J. M. Luttinger and J. C. Ward, Phys. Rev. **118** (1960), 1427.